



ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR - KS09 1336

**PREDIKSI JUMLAH PENDERITA DEMAM
BERDARAH DI KABUPATEN MALANG
BERDASARKAN KONDISI CUACA MENGGUNAKAN
METODE REGRESI DENGAN TRANSFORMASI
NATURAL LOGARITMIK PADA VARIABEL RESPON**

RAFIKA NURMASARI
NRP 5212 100 018

Dosen Pembimbing:
WIWIK ANGGRAENI, S.Si, M.Kom

JURUSAN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2016



ITS
Institut
Teknologi
Sepuluh Nopember

FINAL PROJECT - KS09 1336

**PREDICTING THE NUMBER OF DENGUE PATIENTS
IN MALANG REGENCY BASED ON WEATHER
CONDITION USING REGRESSION
METHODOLOGIES WITH NATURAL LOGARITHM
TRANSFORMED RESPONSE VARIABLES**

RAFIKA NURMASARI
NRP 5212 100 018

Supervisor:
WIWIK ANGGRAENI, S.Si, M.Kom

INFORMATION SYSTEM DEPARTMENT
Faculty of Information Technology
Institut Teknologi Sepuluh Nopember
Surabaya 2016

LEMBAR PENGESAHAN

PREDIKSI JUMLAH PENDERITA DEMAM BERDARAH DI KABUPATEN MALANG BERDASARKAN KONDISI CUACA MENGGUNAKAN METODE REGRESI DENGAN TRANSFORMASI NATURAL LOGARITMIK PADA VARIABEL RESPON

TUGAS AKHIR

Disusun untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Jurusan Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

Oleh:

RAFIKA NURMASARI
NRP 5212100018

Surabaya, April 2016



LEMBAR PERSETUJUAN

PREDIKSI JUMLAH PENDERITA DEMAM BERDARAH DI KABUPATEN MALANG BERDASARKAN KONDISI CUACA MENGGUNAKAN METODE REGRESI DENGAN TRANSFORMASI NATURAL LOGARITMIK PADA VARIABEL RESPON

TUGAS AKHIR

Disusun untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada

Jurusan Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

Oleh:

RAFIKA NURMASARI
NRP 5212100018

Disetujui Tim Penguji:

Tanggal Ujian: 19 April 2016
Periode Wisuda : September 2016

Wiwik Anggraeni, S.Si., M.Kom.


(Pembimbing I)

Mahendrawathi ER., S.T, M.Sc, Ph.D


(Penguji I)

Irmasari Hafidz, S.Kom., M.Sc


(Penguji II)

**LEMBAR PERNYATAAN
PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS**

Sebagai mahasiswa Institut Teknologi Sepuluh Nopember Surabaya, yang bertanda tangan di bawah ini saya :

Nama : RAFIKA NURMASARI
Nrp. : 5212100018
Jurusan / Fak. : SISTEM INFORMASI / FAKULTAS TEKNOLOGI INFORMASI
Alamat kontak :
a. Email : rafikanurmasari@gmail.com
b. Telp/HP : 0821 40004040

Menyatakan bahwa semua data yang saya *upload* di Digital Library ITS merupakan hasil final (revisi terakhir) dari karya ilmiah saya yang sudah disahkan oleh dosen penguji. Apabila dikemudian hari ditemukan ada ketidaksesuaian dengan kenyataan, maka saya bersedia menerima sanksi.

Demi perkembangan ilmu pengetahuan, saya menyetujui untuk memberikan **Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalti-Free Right)** kepada Institut Teknologi Sepuluh Nopember Surabaya atas karya ilmiah saya yang berjudul :

PREDIKSI JUMLAH PENDERITA DEMAM BERDARAH DI KABUPATEN MALANG
BERDASARKAN KONDISI CUACA MENGGUNAKAN METODE REGRESI DENGAN
TRANSFORMASI NATURAL LOGARITMIK PADA VARIABEL RESPON

Dengan Hak Bebas Royalti Non-Eksklusif ini, Institut Teknologi Sepuluh Nopember Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta. Saya bersedia menanggung secara pribadi, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya Ilmiah saya ini tanpa melibatkan pihak Institut Teknologi Sepuluh Nopember Surabaya.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dosen Pembimbing 1

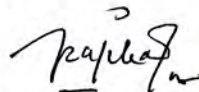
WIWIK ANGGRAENI, S.Si, M.Kom

NIP. 1976 0123 20011 22002

Dibuat di : Surabaya

Pada tanggal : 24 JUNI 2016

Yang menyatakan,



RAFIKA NURMASARI

Nrp. 5212100018

KETERANGAN :

Tanda tangan pembimbing wajib dibubuhi stempel jurusan.

Form dicetak dan diserahkan di bagian Pengadaan saat mengumpulkan hard copy TA/Tesis/Disertasi.

PREDIKSI JUMLAH PENDERITA DEMAM BERDARAH DI KABUPATEN MALANG BERDASARKAN KONDISI CUACA MENGGUNAKAN METODE REGRESI DENGAN TRANSFORMASI NATURAL LOGARITMIK PADA VARIABEL RESPON

Nama Mahasiswa : Rafika Nurmasari
NRP : 5212100018
Jurusan : Sistem Informasi FTIF – ITS
Dosen Pembimbing : Wiwik Anggraeni, S.Si, M.Kom.

ABSTRAK

Penyakit demam berdarah menyebar luas di seluruh daerah tropis dengan variasi risiko lokal yang salah satunya dipengaruhi faktor cuaca seperti curah hujan, temperatur dan kelembaban. Tujuan dari penelitian ini adalah untuk memodelkan hubungan antara faktor cuaca dengan jumlah kasus demam berdarah di Kabupaten Malang, serta memprediksi jumlah kasus demam berdarah berdasarkan faktor cuaca tersebut.

Data yang digunakan dalam penelitian ini adalah data meteorologi dan epidemiologi demam berdarah tahun 2009-2014 per bulan. Data kasus demam berdarah tersedia dari Januari 2009 hingga Desember 2014. Sementara data cuaca tersedia dari Januari 2009 hingga Agustus 2014. Teknik interpolasi dan holt-winters digunakan untuk melengkapi data. Sementara pengujian terhadap berbagai lag faktor cuaca dilakukan untuk mendapatkan variabel cuaca dengan lag yang berkorelasi tertinggi dengan jumlah kasus demam berdarah. Tiga buah model univariat dan sebuah model multivariat

dikembangkan menggunakan metode regresi ordinary least square.

Faktor cuaca di Kabupaten Malang dengan tingkat musiman tertinggi adalah berturut-turut kelembaban, temperatur dan curah hujan. Namun model regresi terbaik adalah yang melibatkan variabel curah hujan dengan lag 7 bulan. Faktor kelembaban dan temperatur tidak lagi mempunyai pengaruh yang signifikan terhadap jumlah kasus demam berdarah ketika faktor curah hujan sudah masuk dalam pertimbangan.

Model yang melibatkan curah hujan sebagai variabel independen merupakan model dengan prediksi terbaik berdasarkan data latih (MAPE 61,07%), sekaligus menjadi yang terburuk berdasarkan data uji (MAPE 79,38%). Model yang menggunakan variabel dependen berupa natural log jumlah kasus demam berdarah menghasilkan prediksi dengan nilai MAPE yang baik, yakni berkisar antara 11,00%-13,67% untuk data latih dan berkisar antara 8,88%-11,89% untuk data uji. Model dengan variabel independen curah hujan kembali menjadi model dengan hasil prediksi terbaik berdasarkan data latih sekaligus terburuk berdasarkan data uji. Dengan demikian regresi OLS cocok digunakan apabila variabel dependen dan independen memiliki varian yang sama, atau setidaknya yang tidak berbeda jauh. Sehingga variabilitas dari variabel dependen dapat dijelaskan dengan baik oleh variabel independen.

Kata kunci: Demam Berdarah Dengue, Cuaca, Regresi Linear Berganda, Ordinary Least Square, Transformasi Natural Log

PREDICTING THE NUMBER OF DENGUE PATIENTS IN MALANG REGENCY BASED ON WEATHER CONDITION USING REGRESSION METHODOLOGIES WITH NATURAL LOGARITHM TRANSFORMED RESPONSE VARIABLES

Name : Rafika Nurmasari
NRP : 5212100018
Department : Sistem Informasi FTIF-ITS
Supervisor : Wiwik Anggraeni, S.Si, M.Kom.

ABSTRACT

Dengue has spreading widely on tropical area with local risk variation one influenced by climatic factors, such as precipitation, temperature and humidity. The objectives of this study is to model the relationship between climatic factors and the number of dengue outbreak in Malang regency, and to predict the number of dengue outbreak based on those climatic factors.

Data used in this study are monthly meteorological and dengue epidemiological from 2009 to 2014. Dengue cases data are available from January 2009 to August 2014. Interpolation and holt-winters were used to complete missing values. Testing against various lag of climatic factors were done to get the optimum lag of climatic variables having the highest correlation with the number of dengue cases. Three models of univariate and a multivariate model were developed using ordinary least square regression method.

The climatic factor in Malang Regency with the highest seasonal level is respectively humidity, temperature and precipitation. Yet the best regression model was the one involving precipitation with a lag of 7 months. Humidity and

temperature are no longer have a significant effect on the number of dengue cases when precipitation has already taken into consideration.

The model with precipitation as the independent variable is the best prediction model based on the trainset (MAPE 61.07%), but being the worst based on testset (MAPE 79.38%). The models using dependent variable in the form of natural log of the number of dengue cases generate predictions with good MAPE value, ranging between 11.00% - 13.67% for the trainset and ranging between 8.88% -11.89% for testset. Model with precipitation as independent variable still has the best predictive results based on trainingset and the worst based on the testset. Thus the OLS regression fit is better used when the dependent and independent variables have the same variant, or at least do not differ much. So that independent variable can explain dependent variable better.

Keyword: Dengue disease, Climate, Multiple Linear Regression, Ordinary Least Squar, Natural Log Transformation

DAFTAR ISI

ABSTRAK	v
KATA PENGANTAR	ix
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
BAB I.....	1
PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Perumusan Masalah.....	4
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	5
1.5. Manfaat Penelitian.....	6
1.6. Relevansi	6
BAB II.....	7
TINJAUAN PUSTAKA.....	7
2.1 Studi Sebelumnya.....	7
2.2 Dasar Teori	11
Data sampel	11
Gambaran umum metode regresi	14
2.2.1. Metode regresi linier berganda	19
2.2.2. Analisis pencilan	20
2.2.3. Analisis multikolinearitas	21
2.2.4. Uji heteroskedastisitas	21
2.2.5. Koefisien determinasi	21
2.2.6. Uji t.....	22
2.2.7. Uji F	23

2.2.8. Kelebihan dan kekurangan metode regresi berganda	24
2.2.9. Evaluasi hasil peramalan	24
2.2.10. Interpolasi	27
2.2.11. <i>Triple Exponential Smoothing</i>	27
BAB III	29
METODOLOGI PENELITIAN	29
3.1. Identifikasi Permasalahan dan Studi Literatur	29
3.2. Pengumpulan dan Pra-proses Data	29
3.3. Membuat Model Regresi	29
3.4. Melakukan Peramalan dan Analisis Hasil.....	31
3.5. Pembuatan Buku Laporan.....	31
BAB IV	33
PERANCANGAN.....	33
4.1. Pengumpulan dan Deskripsi Data	33
4.2. Pra-proses Data.....	33
4.3. Formulasi Model	34
4.4. Analisis Data Awal.....	36
4.5. Pembuatan Model Regresi	37
BAB V	41
IMPLEMENTASI	41
5.1 Pelaksanaan Pra-proses Data	41
5.2 Hipotesis Model Regresi.....	43
5.3 Pembuatan Model H1	47
5.3.1 Model H1a dengan variabel AK.....	47
5.3.2 Model H1b dengan variabel Ln(AK)	54

5.4	Pembuatan Model H2	60
5.4.1	Model H2a dengan variabel AK	60
5.4.2	Model H2b dengan variabel Ln(AK)	67
5.5	Pembuatan Model H3	72
5.5.1	Model H3a dengan variabel AK	72
5.5.2	Model H3b dengan variabel Ln(AK)	79
5.6	Pembuatan Model H4	84
5.6.1	Model H4a dengan variabel AK	84
5.6.2	Model H4b dengan variabel Ln(AK)	88
BAB VI		93
HASIL DAN PEMBAHASAN		93
6.1	Data Hasil	93
6.1.1.	Model H1	93
6.1.2.	Model H2	96
6.1.3.	Model H3	98
6.2	Pembahasan/Diskusi	101
6.3	Usulan Pemilihan/Penggunaan Metode Regresi ..	104
BAB VII		109
KESIMPULAN DAN SARAN		109
7.1	Kesimpulan	109
7.2	Saran	111
Daftar Pustaka		113
BIODATA PENULIS		119
LAMPIRAN A		A-1
LAMPIRAN B		B-1

Halaman ini sengaja dikosongkan.

DAFTAR TABEL

Tabel 2.1 Faktor Lingkungan yang Mempengaruhi Jumlah Kejadian DB.....	11
Tabel 2.2 Interval Nilai Koefisien Korelasi dan Kekuatan Hubungan.....	15
Tabel 2.3 Membaca Nilai t-Test.....	23
Tabel 2.4 Klasifikasi Ukuran Akurasi Hasil Peramalan	25

Tabel 5. 1 Variabel model.....	41
Tabel 5. 2 Evaluasi hasil peramalan variabel cuaca.....	42
Tabel 5. 3 Koefisien pemulusan variabel cuaca.....	42
Tabel 5. 4 Ringkasan numerik variabel	43
Tabel 5. 5 Koefisien korelasi AK dengan variabel cuaca	45
Tabel 5. 6 Rangkuman model regresi H1a	47
Tabel 5. 7 Outlier H1a.....	51
Tabel 5. 8 Perbandingan model H1a lama dan baru	51
Tabel 5. 9 Rangkuman model regresi H1b	54
Tabel 5. 10 Outlier H1b.....	58
Tabel 5. 11 Perbandingan model H1b lama dan baru	59
Tabel 5. 12 Rangkuman model regresi H2a	60
Tabel 5. 13 Outlier H2a.....	63
Tabel 5. 14 Perbandingan model H2a lama dan baru	64
Tabel 5. 15 Rangkuman model regresi H2b	67
Tabel 5. 16 Outlier H2b.....	70
Tabel 5. 17 Perbandingan model H2b lama dan baru	71
Tabel 5. 18 Rangkuman model regresi H3a	72
Tabel 5. 19 Outlier H3a.....	76
Tabel 5. 20 Perbandingan model H3a lama dan baru	76
Tabel 5. 21 Rangkuman model regresi H3b	79
Tabel 5. 22 Outlier H3b.....	83
Tabel 5. 23 Perbandingan model H3b lama dan baru	83
Tabel 5. 24 Rangkuman model regresi H4a	85
Tabel 5. 25 Rangkuman model regresi H4a tereduksi	86

Tabel 5. 26 Rangkuman model regresi H4b	88
Tabel 5. 27 Rangkuman model regresi H4b tereduksi	90

DAFTAR GAMBAR

Gambar 2.1 Pola ICH dan Angka Insiden DBD Kalimantan Timur (AI tinggi) Tahun 2009.....	9
Gambar 2.2 Jumlah Penderita Demam Berdarah Per Bulan ..	12
Gambar 2.3 Temperatur Per Bulan.....	12
Gambar 2.4 Curah Hujan Per Bulan.....	13
Gambar 2.5 Kelembaban Udara per Bulan.....	14
Gambar 2.6 Garis Regresi	15
 Gambar 3. 1. Diagram Alir Pengerjaan Tugas Akhir	30
 Gambar 4. 1 Formulasi kelompok model A	35
Gambar 4. 2 Formulasi kelompok model B.....	36
 Gambar 5. 1 Plot variabel cuaca dan angka kejadian.....	44
Gambar 5. 2 Grafik perbandingan keempat variabel	45
Gambar 5. 3 Plot residual model H1a.....	48
Gambar 5. 4 Plot untuk uji heteroskedastisitas model H1a ...	49
Gambar 5. 5 Plot untuk uji normalitas residual model H1a...	49
Gambar 5.6 Plot untuk analisis pencilan model H1a	50
Gambar 5.7 Plot standardized residual model H1a.....	50
Gambar 5.8 Plot perbandingan heteroskedastisitas model H1a pasca eliminasi pencilan.....	52
Gambar 5.9 Uji normalitas model H1a pasca eliminasi pencilan.....	53
Gambar 5.10 Plot residual model H1b	55
Gambar 5. 11 Plot untuk uji heteroskedastisitas model H1b ..	56
Gambar 5.12 Plot untuk uji normalitas residual model H1b..	56
Gambar 5.13 Plot untuk analisis pencilan model H1b.....	57
Gambar 5.14 Plot standardized residual model H1b.....	58
Gambar 5. 15 Plot residual model H2a.....	61

Gambar 5.16 Plot untuk uji heteroskedastisitas model H2a...	61
Gambar 5. 17 Plot untuk uji normalitas residual	62
Gambar 5. 18 Plot untuk analisis pencilan model H2a	62
Gambar 5. 19 Plot standardized residual model H2a.....	63
Gambar 5. 20 Plot perbandingan heteroskedastisitas model H2a pasca eliminasi pencilan	65
Gambar 5. 21 Uji normalitas model H2a pasca eliminasi pencilan	66
Gambar 5. 22 Plot residual model H2b.....	68
Gambar 5. 23 Plot untuk uji heteroskedastisitas model H2b .	68
Gambar 5. 24 Plot untuk uji normalitas residual model H2b .	69
Gambar 5. 25 Plot untuk analisis pencilan model H2b	69
Gambar 5. 26 Plot standardized residual model H2b.....	70
Gambar 5. 27 Plot residual model H3a.....	73
Gambar 5. 28 Plot untuk uji heteroskedastisitas model H3a..	74
Gambar 5. 29 Plot untuk uji normalitas residual model H3a .	74
Gambar 5. 30 Plot untuk analisis pencilan model H3a	75
Gambar 5. 31 Plot standardized residual model H3a.....	75
Gambar 5. 32 Perbandingan heteroskedastisitas model H3a pasca eliminasi pencilan.....	77
Gambar 5. 33 Uji normalitas model H3a pasca eliminasi pencilan	78
Gambar 5. 34 Plot residual model H3b.....	80
Gambar 5. 35 Plot untuk uji heteroskedastisitas model H3b .	80
Gambar 5. 36 Plot untuk uji normalitas residual model H3b .	81
Gambar 5. 37 Plot untuk analisis pencilan model H3b	82
Gambar 5. 38 Plot standardized residual model H3b.....	82

BAB I

PENDAHULUAN

Pada bagian ini akan dijelaskan mengenai latar belakang masalah, rumusan masalah, batasan masalah dan tujuan penelitian yang mendasari tugas ini.

1.1. Latar Belakang Masalah

Demam berdarah merupakan salah satu penyakit yang banyak ditemui pada daerah beriklim tropis dan sub-tropis, terutama daerah urban dan semi-urban. Menurut WHO (2015) demam berdarah menjadi salah satu penyebab kematian anak-anak terbanyak di beberapa negara Asia dan Amerika Latin [1]. Situs www.healthmap.org menampilkan pemetaan kasus demam berdarah terjadi di seluruh wilayah Indonesia. Setidaknya 21 kabupaten/kota di provinsi Jawa Timur berstatus Kejadian Luar Biasa (KLB) demam berdarah pada Januari 2015. Jumlah penderitanya pun meningkat sebanyak 155,3 persen dari bulan yang sama pada tahun 2014 [2].

Berangkat dari fakta bahwa banyaknya demam berdarah di Indonesia cenderung terjadi ketika musim hujan, maka penggunaan variabel cuaca untuk memprediksi jumlah kejadian demam berdarah menjadi cukup signifikan. Mengutip pernyataan WHO (2015) bahwa penyakit demam berdarah menyebar luas di seluruh daerah tropis, dengan variasi risiko lokal yang dipengaruhi oleh curah hujan, temperatur, dan urbanisasi yang cepat dan tidak terencana [1]. Sumber lain juga menyebutkan bahwa kelembaban udara juga mempengaruhi pola berjangkit infeksi virus dengue. Nyamuk *Aedes* bertahan hidup dalam jangka waktu lama pada suhu panas (28-32°C) dengan kelembaban yang tinggi [3]. Untuk itu, dalam penelitian ini akan dilakukan prediksi jumlah

kejadian demam berdarah di Kabupaten Malang dengan melibatkan faktor cuaca sebagai prediktor.

Penelitian mengenai hubungan cuaca dengan kesehatan manusia masih terus dilakukan karena dinilai berpeluang besar dalam deteksi dini, prediksi, dan pencegahan terhadap efek kesehatan yang merugikan [4]. McLay dkk. (2012) melakukan penelitian mengenai jumlah panggilan medis darurat yang terjadi selama cuaca buruk di kota Hanover, Virginia. Metode yang digunakan adalah metodologi regresi [5]. Disebutkan bahwa cuaca bersalju mempunyai pengaruh yang signifikan di semua model regresi. Variabel lain yang berkorelasi dengan jumlah panggilan darurat adalah kepadatan pada jalan raya yang berkorelasi positif dengan peningkatan jumlah panggilan dan kemungkinan panggilan yang berisiko tinggi. Metode regresi memungkinkan peneliti untuk membuat lebih dari satu model regresi untuk menemukan komposisi variabel yang pas dalam sebuah model.

Model regresi juga pernah dilakukan oleh Andreas Ekstrom dkk. (2014) dalam penelitiannya mengenai prediksi kunjungan departemen darurat sebuah fasilitas kesehatan menggunakan data internet. Hasilnya korelasi antara jumlah kunjungan website antara pukul 6 sore hingga tengah malam adalah signifikan ($r=0,77$; $P<0.001$). Hasil prediksi terbaik dihasilkan dari model yang melibatkan fasilitas kesehatan di seluruh kota, yakni memiliki hasil perhitungan MAPE sebesar 4,8%. Sedangkan untuk model regresi yang hanya melibatkan masing-masing fasilitas kesehatan, memiliki rentang MAPE antara 5,2% hingga 13,1% [6].

Prediksi terhadap munculnya epidemik demam berdarah juga pernah diteliti oleh Elodie Descloux dkk. (2012) dengan studi kasus kota Noumea, New Caledonia. Tujuan dari penelitian mereka adalah untuk memahami hubungan antara variabel-variabel meteorologikal, indeks entomologikal dan dinamika

demam berdarah, mengidentifikasi kondisi ideal terjadinya epidemik demam berdarah, serta membuat model prediktif untuk mendeteksi munculnya demam berdarah. Metode yang digunakan adalah analisis dan pemodelan statistik, antara lain analisis deret waktu (*time series*), bivariat dan multivariat menggunakan regresi linear. Dari sekian banyak model yang dibangun dengan komposisi variabel yang berbeda-beda, didapatkan sebuah model efisien yang dapat menentukan kemunculan penyakit demam berdarah. Model tersebut melibatkan temperatur maksimal dan kelembaban relatif sebagai variabel independen.

Penelitian demam berdarah menggunakan metode *time-series* yang dilakukan oleh Brunkard dkk. (2008) mentransformasi variabel dependennya dengan transformasi natural log. Transformasi tersebut dilakukan untuk menstabilkan varians ketika akan melakukan standar prosedur regresi. Selain varians yang tidak stabil (tidak konstan), transformasi juga dapat digunakan untuk memperbaiki non-linearitas dan residual yang tidak terdistribusi normal (non-normalitas) [27].

Hal menarik dari penggunaan metode statistik ini adalah kemampuannya untuk dapat mengolah potongan-potongan data yang sesuai dengan kriteria tertentu. Misal hanya menggunakan data bulan-bulan tertentu setiap tahunnya, atau melakukan pemilihan tahun-tahun tertentu. Sebagai contoh, penelitian tersebut menemukan bahwa pada bulan Februari di tahun epidemik, curah hujan dan kelembaban relatif mempunyai nilai *t* statistik yang lebih rendah daripada di tahun non-epidemik [7]. Dengan demikian, didapatkan pemahaman yang lebih baik terhadap penyebab-penyebab munculnya demam berdarah oleh faktor-faktor cuaca. Temuan-temuan yang didapatkan dari proses uji statistik tidak dapat digantikan dengan metode-metode lain yang lebih berfokus pada hasil

akhir. Sebagai pembanding, Myriam Gharbi dkk (2011) menggunakan metode SARIMA untuk menghitung kejadian demam berdarah pada tahun 2007 menggunakan data pada tahun 2000 hingga 2007. Hasil yang didapatkan adalah temperatur minimum dengan lag 5 minggu merupakan variabel yang paling baik dalam memprediksi kejadian demam berdarah [8].

Untuk itu, dalam Tugas Akhir ini akan dilakukan prediksi jumlah penderita demam berdarah pada suatu daerah berdasarkan variabel cuaca menggunakan metode regresi. Metode ini cenderung lebih cocok untuk memprediksi variabel dependen yang mempunyai variabel independen yang bersifat kuantitatif. Selain itu metode ini juga bekerja dengan baik ketika lingkup penelitian lebih menyeluruh pada suatu daerah daripada memprediksi masing-masing lingkup yang lebih kecil. Sehingga hasil prediksi jumlah kasus demam berdarah di Kabupaten Malang ini diharapkan dapat membantu fasilitas kesehatan untuk melakukan deteksi dini wabah demam berdarah di daerahnya, sehingga memungkinkan mereka untuk mempersiapkan obat-obatan, tenaga medis, dan aspek lain yang dapat mengurangi dampak kerugian dari demam berdarah.

1.2. Perumusan Masalah

Perumusan masalah yang diangkat pada tugas akhir ini adalah:

1. Model prediksi seperti apa yang dapat menggambarkan jumlah penderita demam berdarah di suatu daerah dengan memperhatikan faktor cuaca pada saat itu?
2. Variabel cuaca yang manakah yang lebih berpengaruh terhadap prediksi jumlah penderita demam berdarah?
3. Bagaimana tingkat signifikansi pengaruh dari variabel-variabel cuaca di dalam model?

1.3. Batasan Masalah

Batasan-batasan dalam pengerjaan tugas akhir ini adalah sebagai berikut:

1. Data penderita demam berdarah yang digunakan adalah penderita demam berdarah yang tercatat di semua puskesmas di Kabupaten Malang dari tahun 2009 sampai 2014.
2. Data cuaca yang digunakan terdiri dari temperatur dan curah hujan bulanan di Kabupaten Malang dari tahun 2009 dan 2014.
3. Metode estimasi regresi linear yang digunakan adalah metode *ordinary least square (OLS)*.
4. Jenis transformasi yang akan digunakan adalah transformasi natural log.

1.4. Tujuan Penelitian

Tujuan tugas akhir ini adalah sebagai berikut:

- a. Memodelkan pengaruh dan hubungan jumlah pasien demam berdarah dengan temperatur serta memprediksi jumlah penderita demam berdarah berdasarkan variabel temperatur tersebut.
- b. Memodelkan pengaruh dan hubungan jumlah pasien demam berdarah dengan kelembaban serta memprediksi jumlah penderita demam berdarah berdasarkan variabel kelembaban tersebut.
- c. Memodelkan pengaruh dan hubungan jumlah pasien demam berdarah dengan curah hujan serta memprediksi jumlah penderita demam berdarah berdasarkan variabel curah hujan tersebut.
- d. Memodelkan pengaruh dan hubungan jumlah pasien demam berdarah dengan kondisi cuaca yang terdiri dari temperatur, kelembaban dan curah hujan serta

memprediksi jumlah penderita demam berdarah berdasarkan tiga variabel tersebut.

1.5. Manfaat Penelitian

Manfaat yang dapat dihasilkan dari tugas akhir ini adalah sebagai berikut:

- a. Bagi penyedia layanan kesehatan dapat memberikan informasi dini tentang prediksi jumlah penderita demam berdarah pada masa yang akan datang sehingga mampu memberikan antisipasi yang baik terhadap lonjakan pasien DB.
- b. Bagi masyarakat umum dapat menjadi peringatan awal datangnya penyakit demam berdarah berdasarkan kondisi cuaca yang dapat teramati.
- c. Bagi akademisi dapat memanfaatkan keilmuannya di bidang ICT untuk diterapkan pada dunia kesehatan.

1.6. Relevansi

Tak lama setelah tiba musim hujan, berita kasus demam berdarah di berbagai daerah di Indonesia mulai membanjiri media pemberitaan lokal hingga nasional. Agar selaras dengan upaya pencegahan wabah demam berdarah, perlu dilakukan prediksi kejadian demam berdarah di masa depan. Hasil temuan pada penelitian ini nantinya dapat dijadikan sebagai dasar melakukan peringatan dini kepada masyarakat pun untuk menentukan waktu yang tepat untuk intervensi wabah demam berdarah.

BAB II

TINJAUAN PUSTAKA

Tinjauan pustaka merupakan penjelasan mengenai teori terkait studi kasus yang disarikan dari buku, jurnal, artikel cetak dan elektronik, atau tugas akhir terdahulu. Berikut ini merupakan penjelasan singkat dari berbagai sumber mengenai metode *Multiple Linear Regression* (regresi linier berganda), penelitian terdahulu, dan teori-teori lain yang digunakan dalam penelitian ini.

2.1 Studi Sebelumnya

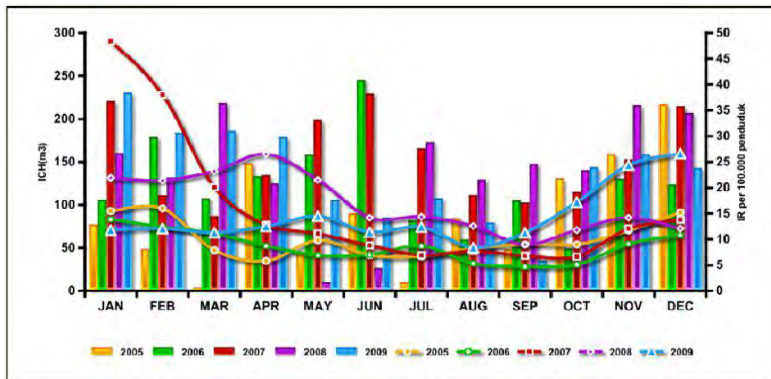
Penelitian yang pernah dilakukan untuk memprediksi suatu nilai variabel berdasarkan variabel lain telah banyak dilakukan. Penelitian lain juga banyak yang mencoba menerapkan konsep korelasi untuk variabel-variabel terkait kesehatan. Berikut adalah hasil dari penelitian sebelumnya:

- a. Andreas Ekstrom dkk. (2014) melakukan penelitian mengenai prediksi kunjungan departemen darurat sebuah fasilitas kesehatan menggunakan data internet. Metode yang digunakan adalah analisis regresi. Hasilnya korelasi antara jumlah kunjungan website antara pukul 6 sore hingga tengah malam adalah signifikan ($r=0,77$; $P<0.001$). Hasil prediksi terbaik dihasilkan dari model yang melibatkan fasilitas kesehatan di seluruh kota, yakni memiliki hasil perhitungan MAPE sebesar 4,8%. Sedangkan untuk model regresi yang hanya melibatkan masing-masing fasilitas kesehatan, memiliki rentang MAPE antara 5,2% hingga 13,1%.

- b. Rogers dkk. (2010) mengadakan penelitian untuk mengidentifikasi peluang untuk meramalkan kesehatan, terutama penyakit-penyakit yang sensitif dengan cuaca. Penelitian ini menyatakan perlunya koordinasi antara sektor kesehatan dengan komunitas pemantau cuaca untuk bersama-sama melakukan program pembangunan dan penelitian. Fokus penelitian ini adalah membuat sebuah sistem peringatan awal (*early warning system*) untuk kesehatan berdasarkan penyebab-penyebab yang berasal dari aspek sosial ekonomi dan lingkungan.
- c. McLay dkk. (2012) melakukan penelitian terhadap pengaruh cuaca buruk terhadap panggilan darurat terkait kesehatan yang diterima oleh *emergency medical service (EMS)*. Fokus penelitian ini adalah memprediksi jumlah panggilan darurat yang tertuju ke EMS selama cuaca buruk agar layanan yang dibutuhkan dapat disiapkan lebih awal. Dengan kata lain, prediksi tersebut bertujuan agar dapat mempersiapkan layanan yang dibutuhkan saat terjadi cuaca buruk [5].
- d. Brunkard dkk (2008) menemukan bahwa faktor-faktor cuaca (temperatur maksimal, temperatur minimal, curah hujan) dan iklim (temperatur permukaan laut) memberikan pengaruh yang kecil namun signifikan dalam insiden demam berdarah di Matamoros, Tamaulipas, Mexico.
- e. Dalam buku pedoman Tata Laksana DBD (2007) yang dikeluarkan oleh Departemen Kesehatan RI, disebutkan bahwa pola berjangkit infeksi virus dengue dipengaruhi oleh iklim dan kelembaban udara. Pada suhu yang panas (28-32°C) dengan kelembaban yang tinggi, nyamuk *Aedes* akan tetap bertahan hidup untuk jangka waktu lama. Selain itu, perbedaan suhu udara dan

kelembaban setiap daerah di Indonesia menyebabkan pola waktu berjangkit di daerah-daerah tersebut juga berbeda. Dicontohkan pula bahwa di Jawa pada umumnya infeksi virus dimulai pada awal Januari dan mencapai puncaknya pada April-Mei setiap tahun.

- f. Topik utama pada Buletin Jendela Epidemiologi (2010) menyebutkan bahwa tersedianya air yang relatif bersih dalam suatu wadah/media akan menjadi tempat perkembangbiakan nyamuk. Telur nyamuk akan menetas setelah 10-12 hari. Dan manusia yang digigit oleh nyamuk dengan virus dengue baru akan menunjukkan gejala infeksi setelah 4-7 hari. Gambar 1 menunjukkan perbandingan pola ICH dengan jumlah insiden per 100.000 penduduk di Kalimantan Timur dari tahun 2005-2009. Sehingga bila hanya memperhatikan faktor risiko curah hujan, waktu yang dibutuhkan dari mulai masuk musim hujan hingga terjadinya insiden DBD adalah sekitar 3 minggu [9].



Gambar 2.1 Pola ICH dan Angka Insiden DBD Kalimantan Timur (AI tinggi) Tahun 2009

- g. Wahyono dkk (2010) melakukan penelitian mengenai hubungan antara faktor individu, faktor lingkungan rumah dan faktor lingkungan luar rumah yang dapat memprediksi terjadinya peningkatan penyakit demam berdarah di kelurahan-kelurahan endemik [10]. Metode yang digunakan terdiri dari 2 tahapan, yaitu 1) assessment untuk melihat gambaran masalah penyakit DB di Kecamatan Cimanggis, 2) studi intervensi untuk melihat efek dari intervensi peningkatan tingkat kepedulian dan partisipasi masyarakat di kelurahan terpilih, dan 3) analisis model spasial untuk melihat kecenderungan dan model untuk melakukan prediksi terhadap kejadian demam berdarah di kecamatan lainnya. Variabel pengamatan yang digunakan dalam penelitian ini antara lain: i) faktor individu (jenis kelamin, umur, lama pendidikan), ii) perilaku (pengetahuan tentang penyebab, tahu penularan DBD, tahu nyamuk *Aedes*, bersihkan bak mandi, bersihkan tempat penampungan air, pakai obat nyamuk, iii) faktor program (penyuluhan dan pemberantasan jentik berkala). Walaupun dalam studi literaturnya disebutkan bahwa faktor-faktor lingkungan fisik seperti terlihat pada Tabel 1 juga mempengaruhi angka kejadian DB, faktor-faktor tersebut tidak ikut menjadi variabel pengamatan dalam penelitiannya.

Tabel 2.1 Faktor Lingkungan yang Mempengaruhi Jumlah Kejadian DB

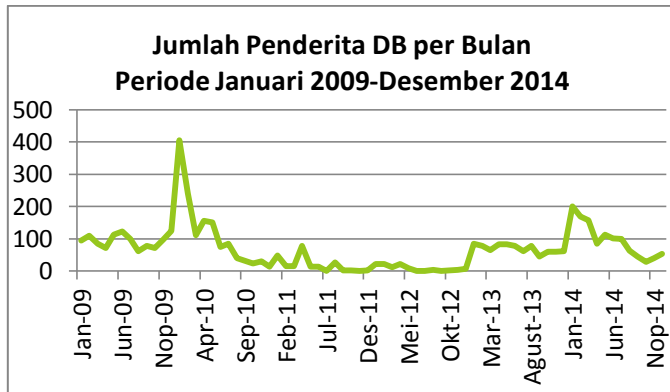
Variabel Dependent	Literatur	Desain Studi	Deskripsi atau Hubungan
Suhu	Pomprou S, 2005	Studi Ekologi	Range 22°C – 33°C
	Yushiro N, et al, 2003	-	Suhu optimum 33°C – 44°C
Kelembaban udara	Pomprou S, 2005	Studi Ekologi	79,19 ± 4,17 %
Hari hujan	Pomprou S, 2005	Studi Ekologi	13,32 ± 6,15 hari

2.2 Dasar Teori

Data sampel

a) Jumlah penderita demam berdarah

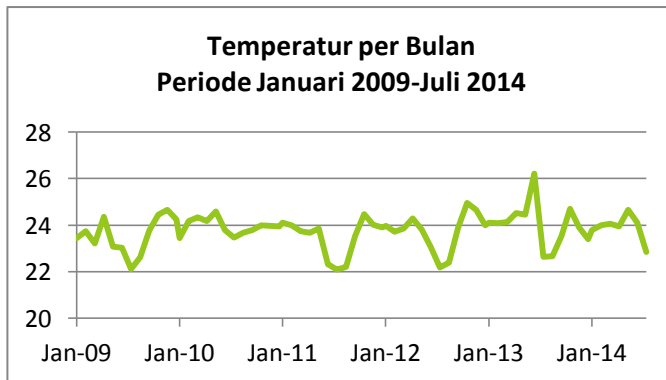
Jumlah penderita demam berdarah merupakan variabel dependen pada tugas akhir ini. Data penderita diambil dari daftar penemuan kasus demam berdarah dari setiap puskesmas di Kabupaten Malang. Kemudian dari daftar tersebut data diagregatkan setiap satu bulan mencakup kasus di semua puskesmas tersebut, ditunjukkan oleh Gambar 2.2.



Gambar 2.2 Jumlah Penderita Demam Berdarah Per Bulan

b) Kondisi cuaca berupa temperatur

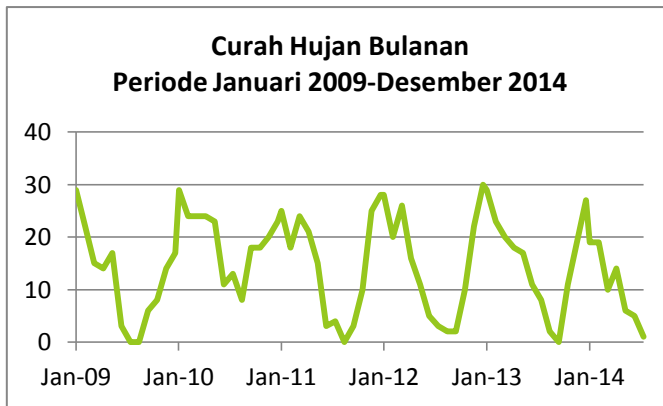
Temperatur diambil dari data cuaca oleh salah satu stasiun pengamat untuk kabupaten Malang. Horizon waktu dari data temperatur, seperti yang nampak pada Gambar 2.3, dipilih per bulan sehingga sama dengan horizon waktu variabel dependen.



Gambar 2.3 Temperatur Per Bulan

c) Kondisi cuaca berupa curah hujan

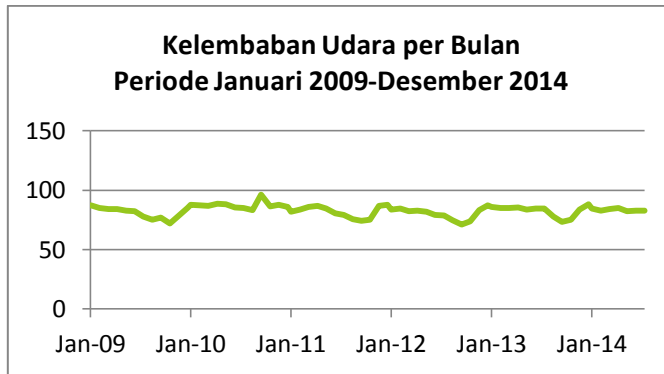
Curah hujan merupakan variabel independen kedua, juga diambil dari data cuaca oleh salah satu stasiun pengamat untuk kabupaten Malang. Horizon waktu dari data curah hujan dipilih per bulan sehingga sama dengan horizon waktu dua variabel lainnya. Gambar 2.4 merupakan grafik curah hujan per bulan di Kabupaten Malang.



Gambar 2.4 Curah Hujan Per Bulan

d) Kelembaban Udara

Variabel independen terakhir yang dijadikan pertimbangan dalam model regresi adalah variabel kelembaban udara. Kelembaban udara di kabupaten Malang selama periode Januari 2009 hingga Desember 2014 seperti Gambar 2.5.



Gambar 2.5 Kelembaban Udara per Bulan

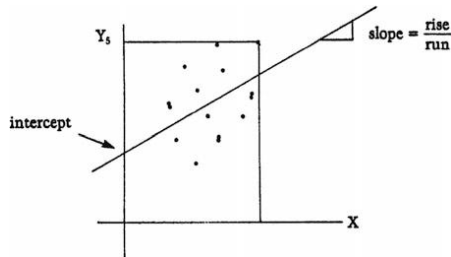
Gambaran umum metode regresi

Metode regresi merupakan salah satu dari model statistik untuk melakukan peramalan, bersama-sama dengan metode exponential smoothing dan beberapa model autoregresif [11]. Metode regresi juga bagian dari model peramalan ekstrapolatif yang mana memiliki asumsi bahwa data pada periode di masa depan akan dipengaruhi oleh variabel yang sama seperti kasus pada periode sebelumnya [12]. Model regresi terdiri dari dua variabel utama, yaitu:

1. Variabel independen adalah variabel yang berdiri sendiri, dalam artian tidak dipengaruhi oleh variabel lain.
2. Variabel dependen adalah variabel yang bergantung pada variabel lain. Dengan kata lain variabel yang nilainya akan ditentukan berdasarkan nilai dari variabel-variabel dependen.

Gambar 2.6 merupakan contoh garis regresi untuk satu variabel dependen Y yang dipengaruhi oleh satu variabel X.

Garis regresi tersebut digambar sedemikian sehingga didapatkan pendekatan terbaik yang menjelaskan hubungan kedua variabel.



Gambar 2.6 Garis Regresi

Ketika garis regresi telah ditentukan, perkiraan nilai variabel independen pada periode peramalan juga harus didapatkan. Sehingga dalam rangka melakukan peramalan yang akurat pada periode tersebut tergantung pada keakuratan estimasi variabel independen.

Analisis regresi biasanya digunakan untuk mengukur bagaimana perubahan yang terjadi pada variabel independen berpengaruh terhadap perubahan variabel dependen. Besarnya kekuatan hubungan antara variabel independen dan variabel dependen tersebut dinyatakan dalam sebuah konstanta yaitu koefisien korelasi (KK). Tabel 2.2 merupakan interval nilai koefisien korelasi yang dapat dijadikan patokan untuk mengukur keeratan sebuah korelasi antar variabel [13].

Tabel 2.2 Interval Nilai Koefisien Korelasi dan Kekuatan Hubungan

No.	Interval Nilai	Kekuatan Hubungan
1	KK = 0,00	Tidak ada

No.	Interval Nilai	Kekuatan Hubungan
2	$0,00 < KK \leq 0,20$	Sangat rendah atau lemah sekali
3	$0,20 < KK \leq 0,40$	rendah atau lemah, tapi pasti
4	$0,40 < KK \leq 0,70$	Cukup berarti atau sedang
5	$0,70 < KK \leq 0,90$	Tinggi atau kuat
6	$0,90 < KK < 1,00$	Sangat tinggi atau kuat sekali, dapat diandalkan
7	$KK = 1,00$	Sempurna

Koefisien korelasi yang digunakan untuk menjelaskan kekuatan hubungan antar variabel bernilai interval/rasio dengan variabel yang bernilai interval/rasio pula adalah koefisien korelasi Pearson (r). Persamaan koefisien korelasi Pearson dapat dihitung dengan persamaan berikut:

$$r = \frac{n \sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{\sqrt{\left[n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2 \right] \left[n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2 \right]}}$$

dimana,

r = koefisien korelasi Pearson

X = variabel independen

Y = variabel dependen

n = banyaknya observasi

Nilai koefisien korelasi dapat bernilai positif atau negatif, yakni berkisar antara -1 dan +1. Nilai positif berarti kedua

variabel berbanding lurus, sedangkan nilai negatif menunjukkan kedua variabel berbanding terbalik.

Setelah mengetahui kekuatan korelasi antar variabel tersebut, kemudian dibentuklah persamaan regresi yang melibatkan variabel-variabel tersebut. Adapun bentuk sederhana dari regresi linear adalah

$$Y = \alpha + \beta X + \varepsilon$$

dimana Y adalah variabel dependen, fungsi linear dari variabel independen X . Parameter α dan β adalah karakteristik garis regresi, dan ε adalah nilai kesalahan acak persamaan regresi. Simpel regresi menghasilkan nilai estimasi nilai Y dalam bentuk:

$$\hat{Y} = a + bX$$

dimana a dan b adalah konstanta regresi yang harus diestimasi. Dengan menggunakan estimasi regresi *least square*, dimana jumlah kuadrat eror diminimalkan dengan menggunakan pendekatan estimasi nilai a dan b sebagai berikut:

1. Menentukan komponen gangguan model regresi.

$$e_i = \hat{Y}_i - Y_i$$

dimana e_i adalah komponen kesalahan estimasi kuadrat terkecil, \hat{Y}_i adalah nilai estimasi variabel dependen dan Y_i adalah nilai aktual variabel dependen. Metode kuadrat terkecil, meminimalkan

$$\sum_{i=1}^N e_i^2 = \sum_{i=1}^N [Y_i - \hat{Y}_i]^2 = \sum_{i=1}^N [Y_i - (a + bX_i)]^2$$

2. Melakukan turunan parsial $\sum_{i=1}^N e_i^2$ terhadap a dan b

$$\begin{aligned}\frac{d \sum_{i=1}^N e_i^2}{d a} &= 2 \sum_{i=1}^N (Y_i - a - bX_i)(-1) \\ &= -2 \left(\sum_{i=1}^N Y_i - \sum_{i=1}^N a - b \sum_{i=1}^N X_i \right) \\ \frac{d \sum_{i=1}^N e_i^2}{d b} &= 2 \sum_{i=1}^N (Y_i - a - bX_i)(-X_i) \\ &= -2 \left(\sum_{i=1}^N Y_i X_i - a \sum_{i=1}^N X_i - b \sum_{i=1}^N X_i^2 \right)\end{aligned}$$

3. Menjadikan kedua persamaan sama dengan nol dan menuliskannya kembali menjadi persamaan normal

$$-2 \left(\sum_{i=1}^N Y_i - \sum_{i=1}^N a - b \sum_{i=1}^N X_i \right) = 0$$

$$\sum_{i=1}^N Y_i = \sum_{i=1}^N a + b \sum_{i=1}^N X_i$$

$$-2 \left(\sum_{i=1}^N Y_i X_i - a \sum_{i=1}^N X_i - b \sum_{i=1}^N X_i^2 \right) = 0$$

$$\sum_{i=1}^N Y_i X_i = a \sum_{i=1}^N X_i + b \sum_{i=1}^N X_i^2$$

4. Menyelesaikan dua persamaan pada poin sebelumnya sehingga didapatkan nilai a dan b

$$\hat{a} = \frac{(\sum_{i=1}^N X_i^2)(\sum_{i=1}^N Y_i) - (\sum_{i=1}^N X_i Y_i)}{N(\sum_{i=1}^N X_i^2) - (\sum_{i=1}^N X_i)^2}$$

$$\hat{b} = \frac{(\sum_{i=1}^N X_i Y_i) - (\sum_{i=1}^N X_i)(\sum_{i=1}^N Y_i)}{N(\sum_{i=1}^N X_i^2) - (\sum_{i=1}^N X_i)^2}$$

2.2.1. Metode regresi linier berganda

Metode regresi linier berganda merupakan salah satu teknik kuantitatif yang dapat digunakan untuk memperkirakan nilai dari sebuah variabel berdasarkan nilai dari variabel-variabel lain yang telah diketahui. Tujuan metode ini adalah menghasilkan sebuah model yang nantinya digunakan untuk mendapatkan nilai variabel dependen dari dua atau lebih variabel independen. Bentuk umum dari model regresi linier berganda dengan variabel dependen y adalah:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

dengan y adalah variabel yang dependen terhadap variabel-variabel x_1 hingga x_n . Dalam notasi matrix, regresi berganda dapat ditulis sebagai berikut:

$$Y = X\beta + \varepsilon$$

Nilai kesalahan pada regresi linear berganda harus memiliki distribusi normal (*bell curve distribution*), yakni mempunyai kemungkinan bernilai nol dan 67,6% nilai kesalahan yang diobservasi memiliki nilai $1 \pm$ standar deviasi nilai kesalahan. Selain itu, sebanyak 95% dari observasi harus berada pada kisaran nilai $2 \pm$ standar deviasi [12]. Secara umum, metode regresi menggunakan OLS (Ordinary Least Square) menggunakan asumsi-asumsi berikut [12] [14]:

- Hubungan antara variabel dependen Y dan variabel independen X adalah linier

- **Normalitas:** variabel gangguan berdistribusi normal dan nilai harapan dari variabel gangguan adalah nol.
- **Homoskedastisitas:** varians dari variabel gangguan adalah sama untuk setiap variabel bebas
- **Tidak ada** korelasi antar variabel gangguan (**autokorelasi**)
- **Kolinearitas:** tidak ada korelasi antar variabel bebas (multikolinearitas)
- Nilai X tetap untuk observasi yang berulang-ulang (non-stokastik).

2.2.2. Analisis pencilan

Pencilan atau yang biasa disebut dengan *outlier* adalah data observasi yang mempunyai nilai ekstrem, yang dapat mengganggu estimasi garis regresi dan residual yang bersangkutan. Terdapat beberapa metode yang dapat digunakan untuk mengidentifikasi *outlier*. Sembilan puluh lima persen tingkat kepercayaan (*confidence interval*) dapat digunakan untuk menghitung berapa banyak data observasi yang berada di luar batas ini. Perbandingan antara residual dibagi dengan standar deviasi residual tersebut (*standardized residual*) haruslah sebuah variabel normal [12].

Jika data observasi i pada suatu regresi dihilangkan, maka $DFBETA_i$ merupakan besar perubahan nilai koefisien regresi dan *studentized residual* adalah nilai residual setelah data observasi i tersebut dihilangkan.

$$DFBETA_i = \frac{(X'X)^{-1}X'e_i}{1 - h_i}$$

dimana $h_i = X_i(X'X)^{-1}X'_i$ dihitung setelah menghapus observasi i , dan

$$sr_i = \frac{e_i}{\hat{\sigma}\sqrt{1-h_i}}$$

dimana $\hat{\sigma}$ adalah standar deviasi dari residual setelah observasi i dihapus. Nilai *studentized residual* (sr) yang lebih dari 2,0 menunjukkan observasi tersebut berpotensi menjadi *outlier*.

2.2.3. Analisis multikolinearitas

Multikolinearitas merupakan kondisi dimana terdapat korelasi antar variabel independen. Adanya multikolinearitas pada sebuah model regresi menyebabkan peningkatan kesalahan standar estimasi seiring dengan penambahan variabel independen. Selain itu kemungkinan mendapatkan hipotesis yang salah juga meningkat [15]. Salah satu cara untuk mendeteksi adanya multikolinearitas adalah dengan melihat matrix kolinear antar variabel.

2.2.4. Uji heteroskedastisitas

Heteroskedastisitas (*heteroscedasticity*) adalah sebuah kondisi dimana residual membentuk sebuah fungsi dari variabel independen. Padahal menurut teori statistik Durbin Watson, nilai residual harus bersifat independen. Sehingga untuk memastikan ada atau tidak ada hubungan antara keduanya, dilakukan pemetaan nilai residual dan variabel independen yang bersangkutan [12].

2.2.5. Koefisien determinasi

Koefisien Determinasi merupakan salah satu metode yang sering digunakan untuk melakukan uji Goodness of FIT, yakni untuk mengukur ketepatan fungsi regresi dalam memprediksi nilai dependen variabel. Koefisien determinasi dilambangkan dengan simbol K_d (R^2) memiliki nilai antara 0 dan 1, berguna untuk mengukur kemampuan model dalam menjelaskan variabel dependen. Semakin besar nilainya maka kemampuan variabel independen dalam menjelaskan semakin baik. Sederhananya, jika $R^2 = a$ artinya semua variabel independen yang ada dalam model dapat menerangkan ($a \cdot 100$) persen variasi dari variabel dependen [14]. Penggunaan koefisien ini perlu diikuti dengan penggunaan nilai *adjusted* R^2 karena koefisien determinasi memiliki kelemahan yaitu bias terhadap jumlah variabel independen yang dimasukkan ke dalam model sehingga nilai R dapat meningkat apabila dilakukan penambahan variabel independen meskipun variabel tersebut tidak terlalu berpengaruh dengan variabel dependen. Sedangkan naik turunnya nilai *adjusted* R^2 sangat dipengaruhi oleh penambahan variabel dependen [15].

2.2.6. Uji t

Uji t digunakan untuk menguji hipotesis nol, dengan maksud untuk memastikan bahwa nilai koefisien a dan b tidak sama dengan nol. H_0 merupakan hipotesis nol, sedangkan H_a merupakan hipotesis alternatif. Berikut adalah tabel cara membaca nilai t statistik dari salah satu sumber:

Tabel 2.3 Membaca Nilai t-Test

Hipotesa satu arah	
Hipotesa positif $H_0 = \text{noI}$ $H_a > \text{noI}$	Hipotesa negatif $H_0 = \text{noI}$ $H_a < \text{noI}$
$t\text{-stat} > t\text{-tabel} : H_0$ ditolak $t\text{-stat} < t\text{-tabel} : H_0$ diterima	$t\text{-stat} < t\text{-tabel} : H_0$ ditolak $t\text{-stat} > t\text{-tabel} : H_0$ diterima
Hipotesa dua arah	
$H_0 = 0$ $H_a \neq 0$	$ t\text{-stat} > t\text{-tabel} : H_0$ ditolak $ t\text{-stat} < t\text{-tabel} : H_0$ diterima

Uji hipotesa satu arah digunakan ketika arah hubungan antar variabel sudah ditentukan. Sedangkan uji hipotesis dilakukan dua arah jika hipotesis belum mempunyai arah atau ingin mengetahui apakah variabel tertentu saling mempengaruhi [16].

2.2.7. Uji F

Uji F digunakan untuk mengetahui apakah semua variabel independen pada model mempunyai pengaruh yang simultan terhadap variabel dependen, dimana

$$H_0 : \alpha = \beta = 0$$

$$H_a : \alpha \neq \beta \neq 0$$

dimana H_0 berarti tidak terdapat pengaruh yang signifikan secara bersama-sama oleh variabel independen terhadap variabel dependen. Sedangkan H_a berarti terdapat pengaruh yang signifikan secara bersama-sama oleh variabel independen terhadap variabel dependen. Jika nilai F-stat melebihi nilai F-tabel pada derajat kebebasan tertentu pada tingkat signifikansi 0,05, maka hipotesis nol ditolak.

2.2.8. Kelebihan dan kekurangan metode regresi berganda

Kelebihan utama dari regresi berganda adalah mudah untuk digunakan karena didukung oleh banyak perangkat lunak komputasi, seperti Microsoft Excel, SAS, SCA, RATS, dan EVIEWS. Sebagai salah satu dari metode statistik, metode regresi juga memiliki kelebihan-kelebihan umum yang dimiliki oleh metode statistik [13], yaitu:

- 1) Memungkinkan mendeskripsikan sesuatu secara eksak
- 2) Dapat menetapkan sampai tingkat mana kesimpulan tersebut benar
- 3) Memberikan rangkuman hasil penelitian dalam bentuk yang lebih berarti dan lebih ringkas
- 4) Dapat menarik kesimpulan umum
- 5) Memungkinkan untuk mengadakan ramalan.

Sedangkan kekurangannya adalah metode ini mengasumsikan bahwa hubungan antara variabel dependen dan independen yang terjadi di masa lalu akan terus terjadi pada masa depan. Sehingga pelaku peramalan harus jeli melihat kondisi tidak umum yang mungkin akan berpengaruh pada variabel dependen [12].

2.2.9. Evaluasi hasil peramalan

Beberapa ukuran ketepatan hasil peramalan yang sering digunakan adalah MSE, MAPE, dan MAD. Makridakis dan Hibon mengklasifikasikan alat ukur ketepatan ramalan seperti yang terlihat pada Tabel 2.4 [17].

Tabel 2.4 Klasifikasi Ukuran Akurasi Hasil Peramalan

Karakteristik pengukuran	Evaluasi dilakukan pada	
	Metode tunggal	Lebih dari satu metode
Absolut	MSE MAE	
Relatif pada metode lain		PB (<i>Percentage Better</i>)
Relatif pada besar kesalahan	MAPE MdAPE	MAPE MdAPE

Berikut merupakan paparan dari beberapa ukuran evaluasi hasil peramalan yang paling populer, yakni MSE, MAE, MAPE, dan MdAPE.

a) Mean Squared Error (MSE)

Mean Squared Error adalah ukuran terpenting yang digunakan untuk mengevaluasi kinerja dari sebuah prediktor atau estimator. MSE digunakan untuk mengukur tingkat akurasi pada pendekatan statistik [18], ditunjukkan oleh formula berikut:

$$MSE = \frac{1}{N} \sum_{k=1}^N (F_k - A_k)^2$$

dimana:

N : jumlah data

F_k : nilai prediksi untuk observasi ke-k

A_k : nilai aktual untuk observasi ke-k

b) Mean Absolute Error (MAE)

Mean Absolute Error merupakan salah satu ukuran absolut seperti MSE, dimana mempunyai nilai 0 hingga $+\infty$. MAE memberikan informasi mengenai rata-rata ukuran kesalahan peramalan ketika nilai negatif diabaikan [17]. Berikut adalah persamaan MAE:

$$MAE = \frac{1}{N} \sum_{k=1}^N |F_k - A_k|$$

c) Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error adalah ukuran kesalahan peramalan yang paling umum, bekerja dengan baik ketika tidak ada nilai ekstrem dalam data (termasuk nol) [19]. MAPE merupakan rata-rata persentase kesalahan absolut dari setiap periode waktu, ditunjukkan oleh formula berikut:

$$MAPE = \frac{1}{N} \sum_{k=1}^N \frac{|F_k - A_k|}{A_k}$$

dimana:

N : jumlah data

F_k : nilai prediksi untuk observasi ke-k

A_k : nilai aktual untuk observasi ke-k

Nilai MAPE yang kecil menunjukkan model peramalan yang cukup baik. Rentang nilai MAPE yang paling baik adalah dibawah 10%, sedangkan untuk nilai MAPE 20%-30% masih dianggap lumrah [20].

d) Median Absolute Percentage Error (MdAPE)

Median Absolute Percentage Error dari sekumpulan data adalah nilai tengah dari persentase kesalahan absolut yang telah diurutkan. Dengan kata lain, MdAPE merupakan nilai median dari sekumpulan nilai APE

(*absoute percentage error*). Nilai APE didapatkan dari persamaan berikut:

$$APE = \frac{|F_k - A_k|}{A_k}$$

dimana:

N : jumlah data

F_k : nilai prediksi untuk observasi ke-k

A_k : nilai aktual untuk observasi ke-k

Setelah nilai APE diurutkan, titik tengah (median) dari APE tersebut dapat dicari dengan persamaan berikut:

$$MdAPE = \frac{N + 1}{2}$$

untuk N ganjil, dan

$$MdAPE = \frac{N}{2} + 1$$

untuk N genap.

Perbedaan MdAPE dengan MAPE adalah kemampuannya untuk menangani adanya *forecast error* yang bernilai ekstrem (*outlier*) [21].

2.2.10. Interpolasi

Interpolasi adalah metode yang digunakan untuk menentukan nilai saat ini atau masa depan sebuah faktor ketika nilai eksak dari faktor tersebut tidak diketahui. Asumsi yang digunakan pada metode ini adalah perubahan antara dua nilai bersifat linear sehingga selisihnya tidak signifikan [22].

2.2.11. Triple Exponential Smoothing

Triple exponential smoothing, atau yang juga dikenal dengan istilah Holt-Winter, merupakan salah satu metode peramalan deret waktu yang memasukkan faktor tren dan musiman [23]. Metode ini mirip dengan moving average, dimana menggunakan data sebelumnya untuk dimasukkan dalam

perhitungan saat ini. Yang membedakan antara keduanya adalah pemberian bobot pada data sebelumnya. Jika moving average memberikan bobot yang sama ke semua data lama, triple exponential smoothing memberikan bobot yang semakin kecil/berkurang secara eksponensial terhadap data yang lebih lama [24]. Adapun persamaan umum dari exponential smoothing adalah:

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1})$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}$$

$$I_t = \beta \frac{y_t}{S_t} + (1 - \beta)I_{t-L}$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m}$$

dimana

y : observasi

t : index mewakili periode waktu

S : pemulusan keseluruhan

b : pemulusan trend

I : pemulusan musiman

F : *forecast*

α , β , γ : konstanta pemulusan

Konstanta pemulusan mempunyai rentang nilai 0-1, ditentukan dengan meminimalkan tingkat kesalahan. Nilai konstanta yang mendekati nol mengindikasikan observasi y yang lebih lama mendapatkan bobot yang lebih besar. Sebaliknya, nilai konstanta yang mendekati satu mengindikasikan observasi yang lebih baru mendapatkan bobot yang lebih besar.

BAB III

METODOLOGI PENELITIAN

Dalam bab ini dijelaskan mengenai tahapan-tahapan yang digunakan untuk menyelesaikan permasalahan tugas akhir ini. Gambar 3.1 merupakan diagram alir pengerjaan, mulai dari identifikasi permasalahan dan studi literatur hingga pembuatan buku laporan akhir. Adapun penjelasan dari masing-masing tahapan tersaji pada bagian sub-bab dari bab ini.

3.1. Identifikasi Permasalahan dan Studi Literatur

Studi literatur dilakukan untuk mengidentifikasi dan memahami permasalahan secara lebih dalam, memilih variabel penelitian yang akan digunakan, serta metode yang akan digunakan dalam penelitian ini.

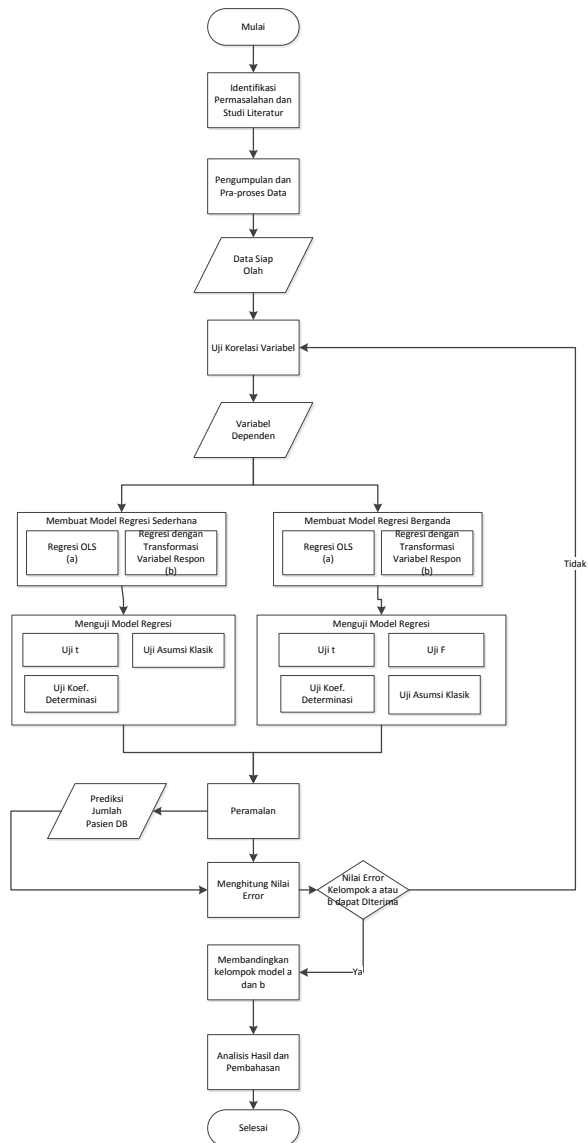
3.2. Pengumpulan dan Pra-proses Data

Data jumlah penderita demam berdarah dan cuaca kabupaten Malang yang telah didapatkan akan melalui tahapan pra-proses data, antara lain memastikan konsistensi data dan menyiapkan data dalam format yang siap untuk diolah. Kemudian data tersebut dibagi dalam dua bagian, yakni data latih dan data uji. Pembagian atau proporsi data yang digunakan untuk data latih dan data uji adalah 75:25.

3.3. Membuat Model Regresi

Pembuatan model regresi untuk studi kasus ini terdiri dari beberapa tahapan, yakni

- i) Uji korelasi variabel untuk mengetahui korelasi antar variabel penelitian, kemudian memilih variabel yang



Gambar 3. 1. Diagram Alir Pengerjaan Tugas Akhir

mempunyai korelasi tinggi terhadap variabel dependen sebagai variabel independen dalam model regresi sederhana dan berganda, baik untuk model dengan variabel respon berupa angka kejadian DBD dan natural log angka kejadian DBD.

- ii) Membuat hipotesis model regresi sederhana dan berganda untuk kelompok model A (model dengan variabel dependen berupa angka kejadian DBD) dan kelompok model B (model dengan variabel dependen berupa natural log angka kejadian DBD).
- iii) Mencari nilai koefisien regresi Ordinary Least Square (OLS) dan regresi dengan transformasi variabel respon natural log untuk membangun model regresi.
- iv) Membuat model regresi menggunakan variabel independen dan nilai koefisien masing-masing variabel tersebut, v) Menguji model regresi dengan koefisien determinasi, uji t, uji asumsi klasik dan uji F (khusus model regresi berganda).

3.4. Melakukan Peramalan dan Analisis Hasil

Setelah persamaan regresi berhasil dibuat, selanjutnya dilakukan peramalan variabel dependen (jumlah pasien demam berdarah) menggunakan model regresi yang sudah dibentuk terhadap data uji. Hasil peramalan yang didapatkan selanjutnya dibandingkan dengan nilai aktual variabel dependen untuk mengetahui tingkat kesalahan hasil peramalan. Apabila dari keJika tingkat kesalahan masih di atas batas toleransi kesalahan, maka kembali ke tahap pembuatan model regresi.

3.5. Pembuatan Buku Laporan

Tahap terakhir dalam penelitian ini adalah pembuatan Buku Laporan, yakni dokumentasi dari setiap tahapan pengerjaan

serta kesimpulan yang didapat dari hasil penelitian dan saran untuk penelitian selanjutnya.

BAB IV

PERANCANGAN

Bab ini berisi rancangan penelitian tugas akhir untuk membuat model peramalan, mulai dari pengumpulan data hingga pembuatan model regresi menggunakan aplikasi Rstudio.

4.1. Pengumpulan dan Deskripsi Data

Baik data jumlah kasus demam berdarah maupun variabel-variabel cuaca adalah data observasional bulanan pada tahun 2009-2014. Data angka kejadian demam berdarah diambil dari daftar penemuan kasus demam berdarah dari setiap puskesmas di Kabupaten Malang dan bersumber dari Dinas Kesehatan Kabupaten Malang. Puskesmas yang dimaksud adalah puskesmas-puskesmas yang tersebar di 39 kecamatan yang berada di kabupaten Malang. Sedangkan data variabel cuaca didapatkan dari stasiun pengamatan cuaca kabupaten Malang. Semua data berupa data kuantitatif dengan unit pengukuran yang berbeda-beda. Unit pengukuran untuk jumlah kasus demam berdarah adalah “kasus”, temperatur dalam derajat celcius, kelembaban dalam “persen”, dan curah hujan dalam “hari hujan”. Khusus untuk variabel curah hujan memiliki nilai nol mutlak yang sah/valid. Untuk variabel lain, nilai nol mengindikasikan *missing value*.

4.2. Pra-proses Data

Data asli (mentah) jumlah kasus demam berdarah di kabupaten Malang yang didapatkan adalah berupa file *excel*. Ringkasan dari data tersebut terdapat pada Lampiran A. Pada beberapa kecamatan, ditemukan adanya beberapa entri kosong

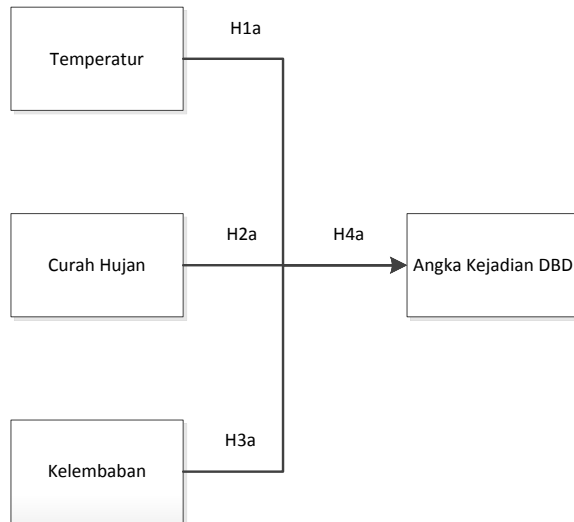
pada tahun 2011 dan 2012. Teknik interpolasi digunakan untuk melengkapi data ini.

Beberapa entri data pada variabel cuaca juga didapati kurang lengkap, tepatnya pada bulan September-Desember 2014. Pendekatan yang digunakan untuk melengkapi data tersebut adalah metode *triple exponential smoothing*. Metode ini dapat menangkap faktor musiman yang menjadi bagian dari variabel cuaca. Data keseluruhan terdiri dari 72 observasi yang dibagi menjadi data latih (*trainset*) dan data uji (*testset*) dengan rasio 75:25.

4.3. Formulasi Model

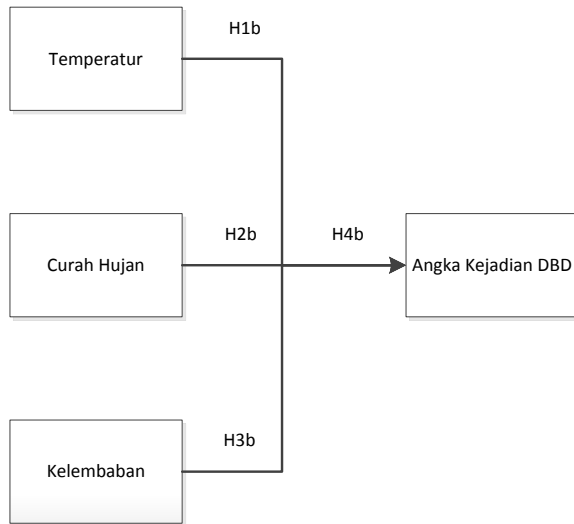
Permasalahan yang diangkat dalam tugas akhir adalah memprediksi jumlah penderita demam berdarah di kabupaten Malang menggunakan variabel-variabel cuaca yang teramati. Model yang akan dibentuk dan diuji coba terdiri dari hubungan-hubungan antar variabel seperti ditunjukkan pada Gambar 4.1 dan Gambar 4.2.

Gambar 4.1 berisi kelompok model A dimana analisis hubungan variabel cuaca dilakukan dengan variabel respon berupa angka kejadian DBD. Sedangkan kelompok model B merupakan model regresi yang menganalisis hubungan antara variabel cuaca dengan angka kejadian DBD dalam bentuk natural log, seperti terdapat pada Gambar 4.2.



Gambar 4. 1 Formulasi kelompok model A

- H1a merupakan model pengaruh variabel temperatur terhadap variabel angka kejadian DBD.
- H2a merupakan model pengaruh variabel curah hujan terhadap variabel angka kejadian DBD.
- H3a merupakan model pengaruh variabel kelembaban terhadap variabel angka kejadian DBD.
- H4a merupakan model pengaruh variabel temperatur, curah hujan, dan kelembaban secara bersama-sama terhadap variabel angka kejadian DBD.



Gambar 4. 2 Formulasi kelompok model B

- H1b merupakan model pengaruh variabel temperatur terhadap variabel natural log angka kejadian DBD.
- H2b merupakan model pengaruh variabel curah hujan terhadap variabel natural log angka kejadian DBD.
- H3b merupakan model pengaruh variabel kelembaban terhadap variabel natural log angka kejadian DBD.
- H4b merupakan model pengaruh variabel temperatur, curah hujan, dan kelembaban secara bersama-sama terhadap variabel natural log angka kejadian DBD.

4.4. Analisis Data Awal

Analisis data awal merupakan langkah penting yang harus selalu dilakukan [25]. Analisa tersebut meliputi:

- Ringkasan numerik: mean, korelasi.
- Ringkasan grafis:
 - Satu variabel: Histogram, diagram garis.

- Dua variabel: Scatter plot.
- Banyak variabel: Grafik interaktif

4.5. Pembuatan Model Regresi

Tahapan-tahapan yang dilakukan untuk mendapatkan sebuah model terbaik untuk masing-masing formulasi model adalah sebagai berikut:

- (1) Menghitung koefisien korelasi terbaik,
yakni menghitung koefisien korelasi antara variabel dependen dengan variabel independen. Variabel independen, yakni variabel-variabel cuaca, pada jeda waktu (*time lag*) tertentu mungkin berkorelasi lebih tinggi dengan variabel dependen daripada variabel independen yang waktu observasinya sama dengan variabel dependen.
- (2) Membuat hipotesis
Hipotesis H1, H2, H3 dan H4 pada kedua kelompok model dibuat berdasarkan pada nilai korelasi yang tertinggi antara variabel dependen dan variabel independen.
- (3) Melakukan regresi
Regresi dilakukan menggunakan aplikasi Rstudio, menghasilkan antara lain persamaan regresi, t-value, p-value, dan koefisien determinasi. Nilai-nilai yang didapatkan disini digunakan untuk tahapan interpretasi hasil dan pengujian asumsi regresi.
- (4) Interpretasi hasil
Derajat kebebasan (*degree of freedom*), t-value. dan p-value digunakan untuk memutuskan penerimaan atau penolakan H0, serta signifikan atau tidaknya pengaruh variabel independen terhadap variabel dependen. Sedangkan Anova (*analysis of variance*) khusus dilakukan pada model regresi berganda.

(5) Uji asumsi regresi

- Uji independensi residual, yakni menguji asumsi residual bersifat independen/bebas dari pengaruh variabel independen dalam model. Dilakukan dengan memetakan nilai residual dengan variabel independen dan menghitung angka korelasinya.
- Uji heteroskedastisitas, yakni menguji asumsi varians dari residual adalah konstan. Dilakukan dengan memetakan nilai residual dengan *fitted value*. Heteroskedastisitas ditandai dengan melebar atau menyempitnya persebaran nilai residual di sekitar garis $y=0$, yang mana berarti bahwa varians tidak konstan. [25] Metode statistik, Breusch-Pagan test, juga digunakan untuk mengecek adanya varians yang tidak konstan. Hipotesis nol dari tes ini adalah bahwa varians dari residual adalah konstan. Nilai $p\text{-value} < .05$ menyebabkan hipotesis nol ditolak, sehingga mengindikasikan adanya heteroskedastisitas. [26]
- Error/residual terdistribusi normal, yakni mengecek distribusi residual untuk memastikan rata-rata residual yang diharapkan bernilai nol dan berdistribusi normal. Dilakukan dengan membuat residual normal QQ plot. Residual yang berdistribusi normal akan tersebar di sekitar garis linear pada plot tersebut.
- Analisis Pencilan (*outlier*), yakni menganalisis kemungkinan terciptanya model yang lebih baik dengan mengeliminasi

observasi yang berpotensi menjadi *outlier*. Dilakukan dengan memanfaatkan plot Jacknife residual (menampilkan index observasi pada plot Q-Q normal dalam bentuk angka), plot standardized residual, plot Cook-statistics (studentized residual). Apabila ditemukan adanya outlier, maka model regresi baru segera dibentuk dan diinterpretasi ulang untuk memastikan apakah model tersebut lebih baik dari model sebelumnya.

Poin 3 sampai 5 dilakukan berulang sampai semua model selesai dibuat. Kemudian model dari kelompok model A dan B dibandingkan berdasarkan variabel independen yang sama.

- (6) Evaluasi hasil peramalan pada data latih.
Alat ukur tingkat kesalahan hasil peramalan yang akan digunakan adalah MAPE.
- (7) Melakukan peramalan dan evaluasi hasil peramalan.
Peramalan dilakukan menggunakan model terbaik pada data uji yang telah disiapkan serta melakukan evaluasi hasil peramalan.

Halaman ini sengaja dikosongkan.

BAB V

IMPLEMENTASI

Bab ini berisi proses pelaksanaan penelitian, yakni tahapan-tahapan dalam pembuatan model, uji asumsi model, hingga penentuan model terbaik.

5.1 Pelaksanaan Pra-proses Data

Variabel-variabel yang akan dilibatkan pada kedelapan model ditunjukkan oleh Tabel 5.1.

Tabel 5. 1 Variabel model

Nama Variabel	Definisi
AK	Angka kejadian demam berdarah, atau jumlah penderita demam berdarah.
Ln AK	Bentuk natural logaritmik dari variabel AK
Temperatur	Temperatur atau suhu udara rata-rata per bulan
Curah.Hujan	Jumlah hari hujan dalam sebulan
Kelembaban	Kelembaban udara rata-rata per bulan

Data hasil pra-proses menggunakan interpolasi terdapat pada Lampiran A: Data hasil pra-proses menggunakan interpolasi. Tabel 5.2 menunjukkan evaluasi peramalan variabel cuaca yang dimaksudkan untuk melengkapi beberapa entri kosong pada variabel-variabel tersebut. Sementara itu, tabel lengkap yang berisi semua data yang digunakan dalam penelitian ini terdapat pada Lampiran B.

Tabel 5. 2 Evaluasi hasil peramalan variabel cuaca

Variabel	<i>MSE</i>	<i>MAE</i>
Temperatur	0,295693588	0,384937699
Curah.Hujan	163,6962025	7,620253165
Kelembaban	13,02816447	2,556163998

Rata-rata kesalahan absolut pada variabel temperatur adalah 0,385. Nilai tersebut cenderung sangat baik (1,62%) karena rata-rata nilai aktual dari variabel temperatur adalah 23,78. Rata-rata kesalahan absolut pada variabel kelembaban adalah 2,556. Nilai tersebut cenderung baik (3,08%) karena rata-rata nilai aktual dari variabel kelembaban adalah 82,82. Sedangkan rata-rata kesalahan absolut pada variabel curah hujan adalah 7,620. Nilai tersebut cenderung tidak begitu baik (48,47%) karena rata-rata nilai aktual dari variabel curah hujan adalah 15,72.

Tabel 5. 3 Koefisien pemulusan variabel cuaca

Variabel	Parameter Holt-Winters	
	beta (tren)	gamma (musiman)
Temperatur	0	0,097115
Curah Hujan	0	0,373024
Kelembaban	0	0,088311

Variabel temperatur dan kelembaban mempunyai nilai konstanta musiman yang hampir mendekati nol. Artinya, observasi yang lebih lama mendapatkan bobot yang lebih besar. Dengan kata lain faktor musimannya sangat kuat. Sementara faktor musiman variabel curah hujan adalah yang paling besar dari dua variabel lainnya. Karena rentang musim yang digunakan untuk semua variabel adalah 12 bulan, hal ini berarti curah hujan tidak sebaik dua variabel lainnya dalam

mengikuti pola musiman 12 bulan tersebut. Sementara faktor tren tidak ditemukan pada ketiga variabel cuaca tersebut.

5.2 Hipotesis Model Regresi

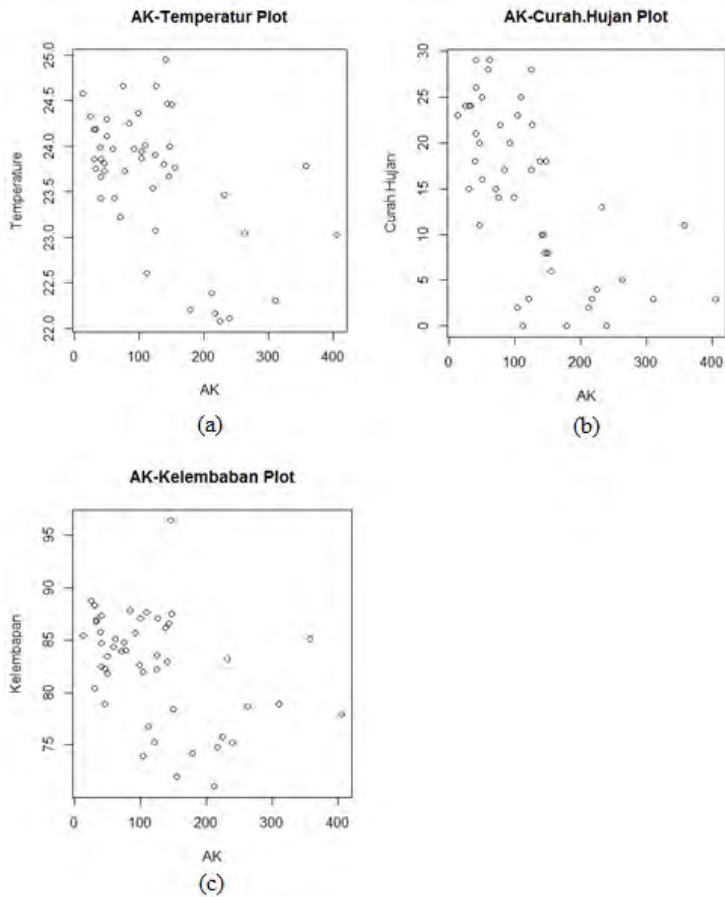
Ringkasan numerik variabel AK dan tiga variabel cuaca lainnya ditunjukkan oleh Tabel 5.3. Ringkasan data semacam ini dimaksudkan untuk melihat data, antara lain memastikan data berada pada rentang nilai yang valid.

Tabel 5. 4 Ringkasan numerik variabel

Ringkasan	Angka kejadian	Temperatur	Curah hujan	Kelembaban
Min.	13,00	22,08	0,00	71,02
Q1	52,25	23,48	8,50	79,30
Median	104,00	23,89	17,00	84,00
Mean	118,48	23,78	15,72	82,82
Q3	144,50	24,18	23,00	85,78
Maks.	405,00	26,22	30,00	96,40

Variabel angka kejadian berada pada rentang nilai 13-405. Variabel temperatur mempunyai rentang nilai 22,08-26,22 dengan rata-rata 23,78. Temperatur tersebut menunjukkan suhu Kabupaten Malang yang cenderung dingin. Variabel curah hujan mempunyai rentang nilai 0-30 dan rata-rata 15,72. Nilai 0 (nol) pada variabel curah hujan bukanlah sebuah *missing value*, seperti yang telah dijelaskan pada bagian deskripsi data. Variabel kelembaban memiliki rentang nilai 71,02-96,40 dan rata-rata 82,82. Di antara tiga variabel cuaca tersebut curah hujan merupakan variabel yang memiliki rentang nilai paling lebar. Plot antara variabel angka kejadian dengan masing-masing variabel cuaca ditunjukkan oleh Gambar 5.1. Ketiga plot tersebut menunjukkan pola hubungan

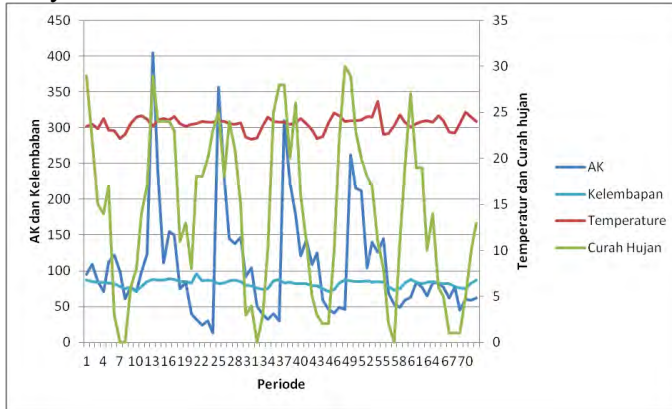
negatif antara variabel angka kejadian berturut-turut dengan variabel temperatur, curah hujan dan kelembaban.



Gambar 5. 1 Plot variabel cuaca dan angka kejadian

Gambar 5.2 menunjukkan grafik nilai variabel AK dan tiga variabel cuaca. Dapat dilihat bahwa keempatnya mengalami gejala musiman yang berbeda-beda, lebih tepatnya periode

terjadinya nilai tertinggi dari setiap variabel. Oleh karena itu untuk menyesuaikan dengan asumsi regresi, dimana hubungan variabel dependen dan independen adalah linier, periode naik turunnya kedua variabel tersebut akan disamakan.



Gambar 5. 2 Grafik perbandingan keempat variabel

Koefisien korelasi *pearson* yang dilakukan pada penelitian ini diterapkan untuk mencari pergeseran setiap variabel cuaca sehingga dapat mencapai maksud yang telah dipaparkan sebelumnya. Hasil koefisien korelasi pada setiap pergeseran (*time lag*) yang mungkin, ditunjukkan pada Tabel 5.4.

Tabel 5. 5 Koefisien korelasi AK dengan variabel cuaca

Time lag	Koefisien korelasi		
	AK-temperatur	AK-curah.hujan	AK-kelembaban
0	0,08187231	0,43261614	0,23496221
1	0.2361706	0.5294655	0.314437
2	0.4356448	0.5075219	0.2564617
3	0.5284899	0.2595873	-0.1083503
4	0.3292826	0.0132155	-0.1655791
5	-0.1172318	-0.370127	-0.3879241

Time lag	Koefisien korelasi		
	AK-temperatur	AK-curah.hujan	AK-kelembaban
6	-0.5114382	-0.5227894	-0.4079432
7	-0.532115	-0.6577457	-0.311765
8	-0.2858383	-0.4198862	-0.1317989
9	-0.07837724	-0.2580046	-0.0266368
10	-0.08354886	-0.01594525	0.10764
11	-0.02472721	0.1558552	0.1822757
12	0.01101504	0.5483105	0.2833429

Semua koefisien korelasi tertinggi pada setiap pasangan variabel bebas dan terikat, seperti ditunjukkan pada Tabel 5.4, memiliki kekuatan hubungan sedang ($0,40 < |\text{koef.korelasi}| \leq 0,70$). Koefisien korelasi tertinggi antara variabel AK dengan variabel Temperatur, Curah.Hujan dan Kelembaban berturut-turut adalah -0,532115 (lag 7), -0,6577457(lag 7), dan -0,4079432 (lag 6). Sehingga hipotesis untuk setiap model regresi yang akan dibuat adalah sebagai berikut:

(1) Hipotesis model H1

H_0 : tidak ada hubungan negatif antara AK dan Temperatur

H_A : ada hubungan negatif antara AK dan Temperatur

(2) Hipotesis model H2

H_0 : tidak ada hubungan negatif antara AK dan Curah.Hujan

H_A : ada hubungan negatif antara AK dan Curah.Hujan

(3) Hipotesis model H3

H_0 : tidak ada hubungan negatif antara AK dan kelembaban

H_A : ada hubungan negatif antara AK dan kelembaban

(4) Hipotesis model H4

H_0 : tidak ada pengaruh yang simultan dari variabel Temperatur, Curah Hujan, dan Kelembaban terhadap variabel AK.

H_A : terdapat pengaruh yang simultan dari variabel Temperatur, Curah Hujan, dan Kelembaban terhadap variabel AK.

5.3 Pembuatan Model H1

5.3.1 Model H1a dengan variabel AK

Berikut ini adalah tahap-tahap pembuatan model regresi untuk variabel AK dan Temperatur:

(1) Melakukan regresi

Persamaan regresi yang dihasilkan adalah:

$$AK = -65,90 * \text{Temperatur} + 1681,50$$

dengan detail yang terangkum dalam Tabel 5.5 di bawah ini.

Tabel 5. 6 Rangkuman model regresi H1a

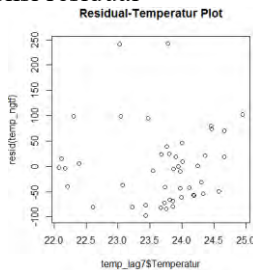
<i>Summary</i>	Korelasi negatif AK-temperatur lag 7
Intercept	1681,50
Slope	-65,90
t value	-4,216
p value	0,0001183
Signif. code	***0,001
Residual std. error	77,01
DF	45
Multiple R-squared	0,2831
Adjusted R-squared	0,2672
F-statistic	17,77

(2) Interpretasi Hasil

- t-tabel dengan degree of freedom 45 dan tingkat signifikansi 0,05 adalah 1.67943
- t-value persamaan regresi adalah -4,216. Karena $t\text{-value} < t\text{-tabel}$, maka H_0 ditolak
- p-value = 0,0001183 menunjukkan pengaruh yang signifikan pada level 0,001.

(3) Uji asumsi regresi

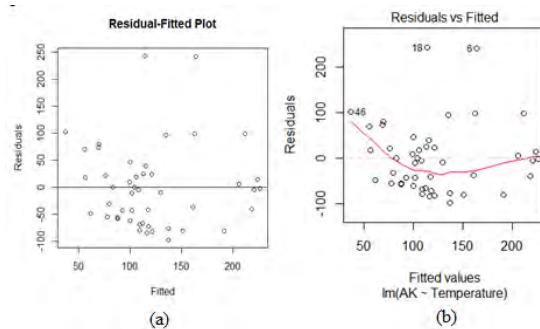
- Uji independensi residual



Gambar 5. 3 Plot residual model H1a

Plot residual pada Gambar 5.3 terdistribusi acak, tidak terlihat adanya sebuah fungsi yang menghubungkan residual dengan variabel Temperatur. Selain itu korelasi antara residual dengan Temperatur adalah $8.148309e^{-16}$. Hal ini menunjukkan residual independen / bebas dari pengaruh variabel Temperatur.

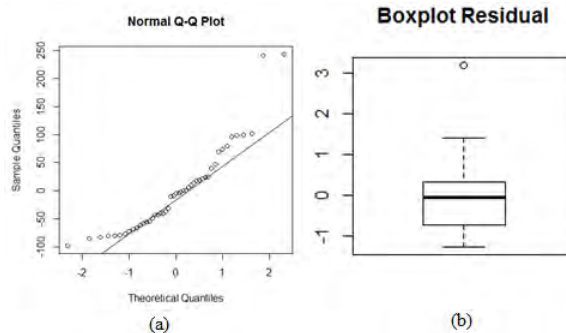
- Uji heteroskedastisitas



Gambar 5. 4 Plot untuk uji heteroskedastisitas model H1a

Pada Gambar 5.4, terlihat kecenderungan data melebar. Namun nilai p-value yang dihasilkan dari tes statistik adalah 0,5030408 sehingga mengindikasikan varians yang telah konstan.

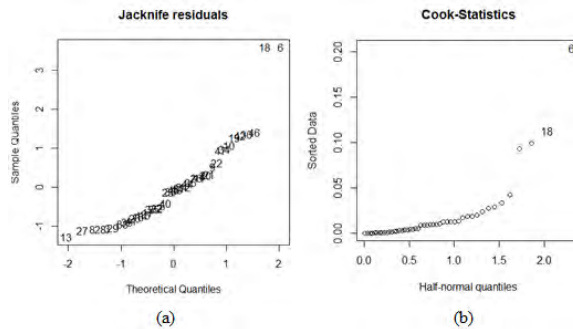
- Residual terdistribusi normal



Gambar 5. 5 Plot untuk uji normalitas residual model H1a

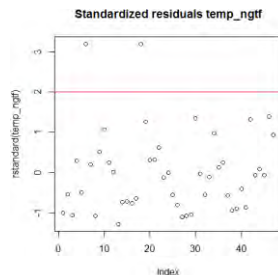
Residual terdistribusi normal karena berada di sekitar garis linear pada normal Q-Q plot (Gambar 5.5 (a)) dan memiliki rata-rata nol pada boxplot residual (Gambar 5.5 (b)).

- Analisis pencilan



Gambar 5.6 Plot untuk analisis pencilan model H1a

Plot Jackknife residual (Gambar 5.6 (a)) menampilkan index observasi pada plot Q-Q normal dalam bentuk angka. Grafik Cook-Statistics (Gambar 5.6 (b)) menunjukkan observasi-observasi yang berpotensi menjadi pencilan dengan menganalisis nilai studentized residual-nya. Observasi yang mempunyai nilai studentized residual lebih dari 2,0 adalah observasi 6 dan 18.



Gambar 5.7 Plot standardized residual model H1a

Standardized residual juga dapat digunakan untuk mendeteksi pencilan. Pencilan pada plot

- standardized residual (Gambar 5.7) adalah observasi yang berada di luar rentang nilai $[-2,2]$.
- Analisis pasca eliminasi pencilan
Observasi yang menjadi pencilan ditunjukkan oleh Tabel 5.6.

Tabel 5. 7 Outlier H1a

Index Obs.	AK	Temperatur
6	405	23,02333
18	357	23,78333

Persamaan regresi baru yang dihasilkan adalah:

$$AK = -60,22 * \text{Temperatur} + 1536,15$$

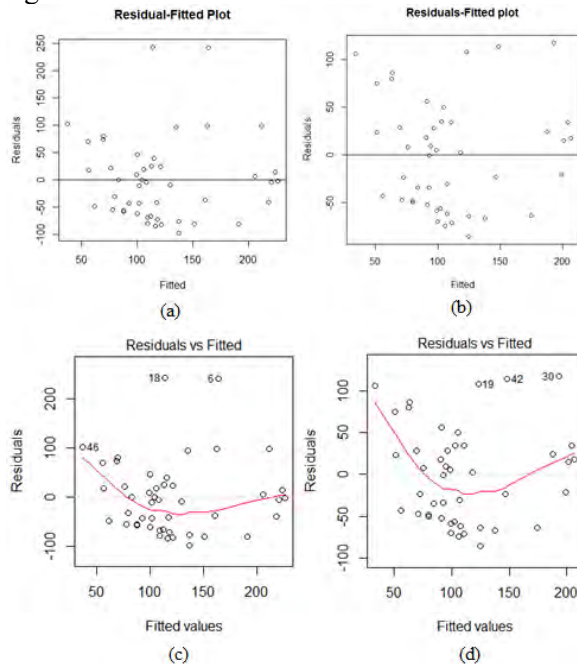
dengan detail perbandingan antara persamaan lama dan baru (setelah eliminasi pencilan) terangkum dalam Tabel 5.7.

Tabel 5. 8 Perbandingan model H1a lama dan baru

<i>Summary</i>	Korelasi negatif AK-temperatur lag 7	Sesudah eliminasi pencilan
Intercept	1681,50	1536,15
Slope	-65,90	-60,22
t value	-4,216	-5,081
p value	0,0001183	$7,76e^{-06}$
Signif. code	***0,001	***0,001
Residual std. error	77,01	57,86
DF	45	43
Multiple R-squared	0,2831	0,3751

<i>Summary</i>	Korelasi negatif AK- temperatur lag 7	Sesudah eliminasi pencilan
Adjusted R-squared	0,2672	0,3606
F-statistic	17,77	25,82

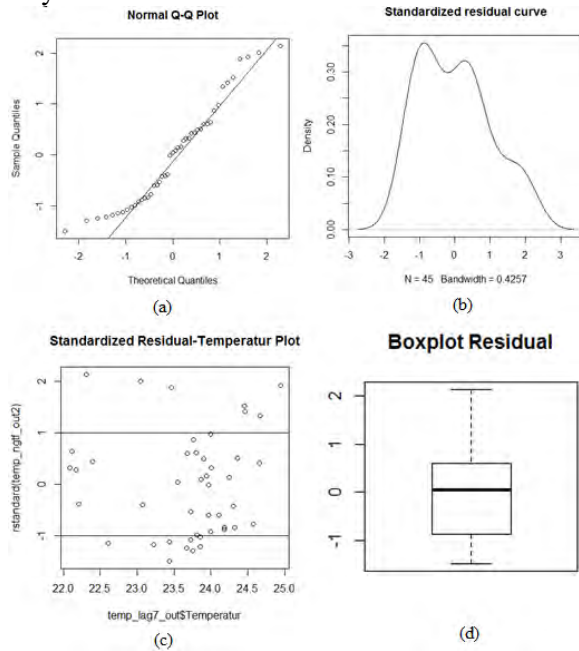
Nilai t-value, p-value, dan R^2 berubah semakin baik. Selanjutnya menguji kembali asumsi regresi.



Gambar 5.8 Plot perbandingan heteroskedastisitas model H1a pasca eliminasi pencilan

Plot pada Gambar 5.8 (a) dan (c) merupakan kepunyaan model sebelum pengeliminasian

pencilan. Setelah pencilan dibuang, nilai p-value meningkat menjadi 0,9752848. Sehingga varians dinyatakan telah konstan.



Gambar 5.9 Uji normalitas model H1a pasca eliminasi pencilan

Pada Gambar 5.9 (c) terdapat 15 dari 45 observasi yang berada diluar rentang nilai $[-1, 1]$. Artinya, 66,67% observasi berada pada rentang nilai tersebut. Persentase tersebut lebih kecil dari 67,6%. Namun, boxplot dari standard residual (Gambar 5.9 (d)) menunjukkan rata-rata residual adalah nol. Sehingga bisa dikatakan model yang baru sudah terdistribusi normal.

(4) Evaluasi hasil peramalan pada data latih

Evaluasi hasil peramalan pada data latih menunjukkan nilai MAPE yang sangat besar, yakni 75,41%.

- (5) Melakukan peramalan dan evaluasi hasil peramalan
Evaluasi hasil peramalan pada data uji menunjukkan nilai MAPE sebesar 61,22%

5.3.2 Model H1b dengan variabel Ln(AK)

Berikut ini adalah tahap-tahap pembuatan model regresi untuk variabel Ln(AK) dan Temperatur:

- (1) Melakukan regresi
Persamaan regresi yang dihasilkan adalah:
$$\text{Ln(AK)} = -0,5412 * \text{Temperatur} + 17,33505$$

dengan detail yang terangkum dalam Tabel 5.8.

Tabel 5. 9 Rangkuman model regresi H1b

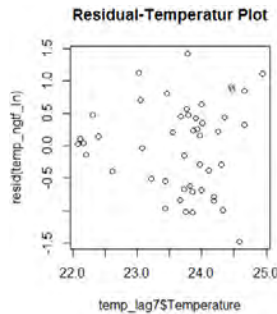
<i>Summary</i>	Korelasi negatif Ln(AK) - temperatur lag 7
Intercept	17,3350
Slope	-0,5412
t value	-3,881
p value	0,000337
Signif. code	***0,001
Residual std. error	0,6869
DF	45
Multiple R-squared	0,2508
Adjusted R-squared	0,2342
F-statistic	15,07

- (2) Interpretasi hasil
- t-tabel dengan degree of freedom 45 dan tingkat signifikansi 0,05 adalah 1.67943

- t-value persamaan regresi adalah -3,881. Karena $t\text{-value} < t\text{-tabel}$, maka H_0 ditolak
- p-value = 0,000337 menunjukkan pengaruh yang signifikan pada level 0,001.

(3) Uji asumsi regresi

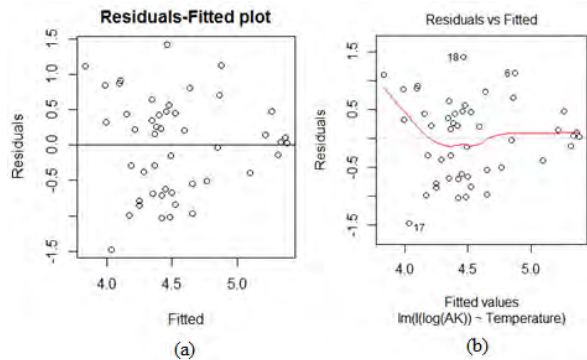
- Uji independensi residual



Gambar 5.10 Plot residual model H1b

Plot residual pada Gambar 5.10 terdistribusi acak, tidak terlihat adanya sebuah fungsi yang menghubungkan residual dengan variabel Temperatur.

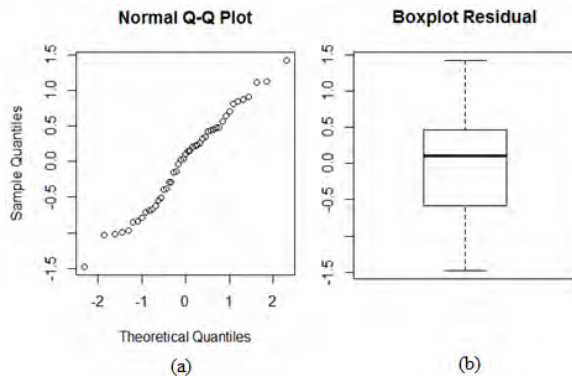
- Uji heteroskedastisitas



Gambar 5. 11 Plot untuk uji heteroskedastisitas model H1b

Pada Gambar 5.11, terlihat kecenderungan data menyempit. Namun nilai p-value dari tes statistik adalah 0,05441371, sehingga mengindikasikan varians telah konstan.

- Residual terdistribusi normal

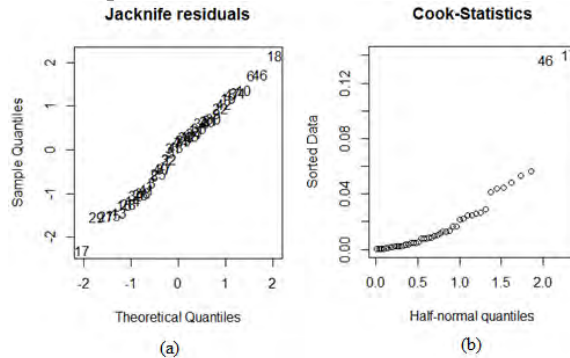


Gambar 5.12 Plot untuk uji normalitas residual model H1b

Residual terdistribusi normal karena berada di sekitar garis linear pada normal Q-Q plot

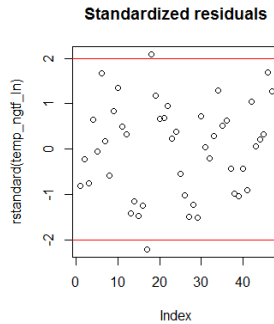
(Gambar 5.12 (a)) dan memiliki rata-rata nol pada boxplot residual (Gambar 5.12 (b)).

– Analisis pencilan



Gambar 5.13 Plot untuk analisis pencilan model H1b

Plot Jackknife residual (Gambar 5.13 (a)) menampilkan index observasi pada plot Q-Q normal dalam bentuk angka. Grafik Cook-Statistics (Gambar 5.13 (b)) menunjukkan observasi-observasi yang berpotensi menjadi pencilan dengan menganalisis nilai studentized residual-nya. Observasi yang mempunyai nilai studentized residual lebih dari 2,0 adalah observasi 46 dan 17.



Gambar 5.14 Plot standardized residual model H1b

- Standardized residual juga dapat digunakan untuk mendeteksi pencilan. Pencilan pada plot standardized residual (Gambar 5.14) adalah observasi yang berada di luar rentang nilai $[-2, 2]$
- Analisis pasca eliminasi pencilan
Observasi yang menjadi pencilan ditunjukkan oleh Tabel 5.9.

Tabel 5. 10 Outlier H1b

Index Obs.	Ln(AK)	Temperatur
17	2,564949	24,57742
46	4,941642	24,94435

Persamaan regresi baru yang dihasilkan adalah:

$$\text{Ln(AK)} = -0,5440 * \text{Temperatur} + 17,4089$$

dengan detail perbandingan antara persamaan lama dan baru (setelah eliminasi pencilan) terangkum dalam Tabel 5.10.

Tabel 5. 11 Perbandingan model H1b lama dan baru

<i>Summary</i>	Korelasi negatif Ln(AK)- temperatur lag 7	Sesudah eliminasi pencilan
Intercept	17,3350	17,4089
Slope	-0,5412	-0,5440
t value	-3,881	-3,936
p value	0,000337	0,000298
Signif. code	***0,001	***0,001
Residual std. error	0,6869	0,6443
DF	45	43
Multiple R- squared	0,2508	0,2649
Adjusted R- squared	0,2342	0,2478
F-statistic	15,07	15,49

Pengeliminasian pencilan pada model tidak berpengaruh banyak terhadap perubahan nilai intercept dan slope. Sehingga kedua pencilan tersebut dapat dibiarkan.

- (4) Evaluasi hasil peramalan pada data latih
Evaluasi hasil peramalan pada data latih menunjukkan nilai MAPE 13,60%.
- (5) Melakukan peramalan dan evaluasi hasil peramalan
Evaluasi hasil peramalan pada data uji menunjukkan nilai MAPE 9,02%.

5.4 Pembuatan Model H2

5.4.1 Model H2a dengan variabel AK

Berikut ini adalah tahap-tahap pembuatan model regresi untuk variabel AK dan Curah.Hujan:

- (1) Melakukan regresi

Persamaan regresi yang dihasilkan adalah:

$$AK = -6,511 * \text{curah hujan} + 218,493$$

dengan detail yang terangkum dalam Tabel 5.11.

Tabel 5. 12 Rangkuman model regresi H2a

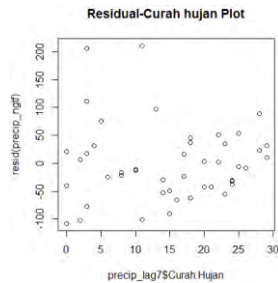
<i>Summary</i>	Korelasi negatif AK-curah hujan lag 7
Intercept	218,493
Slope	-6,511
t value	-5,858
p value	$5,08e^{-07}$
Signif. code	***0,001
Residual std. error	68,52
DF	45
Multiple R-squared	0,4326
Adjusted R-squared	0,42
F-statistic	34,31

- (2) Interpretasi Hasil

- t-tabel dengan degree of freedom 45 dan tingkat signifikansi 0,05 adalah 1.67943
- t-value persamaan regresi adalah -5,858. Karena $t\text{-value} < t\text{-tabel}$, maka H_0 ditolak
- $p\text{-value} = 5,08e^{-07}$ menunjukkan pengaruh yang signifikan pada level 0,001.

- (3) Uji asumsi regresi

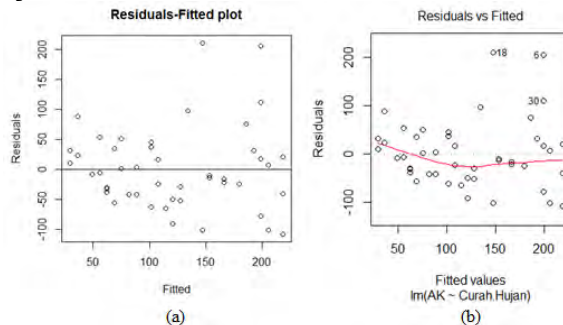
- Uji independensi residual



Gambar 5. 15 Plot residual model H2a

Plot residual pada Gambar 5.15 terdistribusi acak, tidak terlihat adanya sebuah fungsi yang menghubungkan residual dengan variabel Curah.Hujan. Selain itu korelasi antara residual dengan Curah.Hujan adalah $6.775961e^{-17}$. Hal ini menunjukkan residual independen/bebas dari pengaruh variabel Curah.Hujan.

- Uji heteroskedastisitas

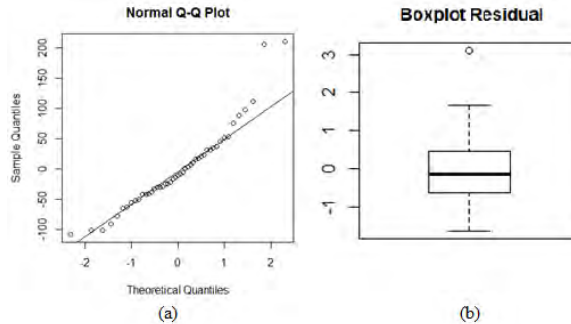


Gambar 5.16 Plot untuk uji heteroskedastisitas model H2a

Pada Gambar 5.16, terlihat kecenderungan data melebar. Nilai p-value dari tes statistik adalah

0,005761691 sehingga mengindikasikan varians yang belum konstan.

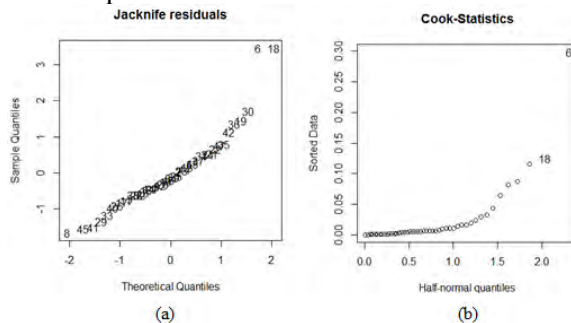
- Residual terdistribusi normal



Gambar 5. 17 Plot untuk uji normalitas residual

Residual terdistribusi normal karena berada di sekitar garis linear pada normal Q-Q plot (Gambar 5.17 (a)) dan memiliki rata-rata nol pada boxplot residual (Gambar 5.17 (b)).

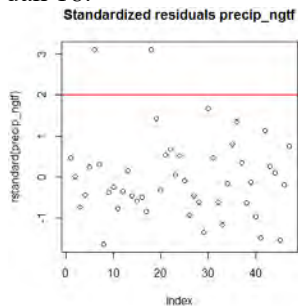
- Analisis pencilan



Gambar 5. 18 Plot untuk analisis pencilan model H2a

Plot Jackknife residual (Gambar 5.18 (a)) menampilkan index observasi pada plot Q-Q

normal dalam bentuk angka. Grafik Cook-Statistics (Gambar 5.18 (b)) menunjukkan observasi-observasi yang berpotensi menjadi pencilan dengan menganalisis nilai studentized residual-nya. Observasi yang mempunyai nilai studentized residual lebih dari 2,0 adalah observasi 6 dan 18.



Gambar 5. 19 Plot standardized residual model H2a

Standardized residual juga dapat digunakan untuk mendeteksi pencilan. Pencilan pada plot standardized residual (Gambar 5.19) adalah observasi yang berada di luar rentang nilai $[-2,2]$.

- Analisis pasca eliminasi pencilan
Observasi yang menjadi pencilan ditunjukkan oleh Tabel 5.12.

Tabel 5. 13 Outlier H2a

Index Obs.	AK	Temperatur
6	405	3
18	357	11

Persamaan regresi yang dihasilkan adalah:

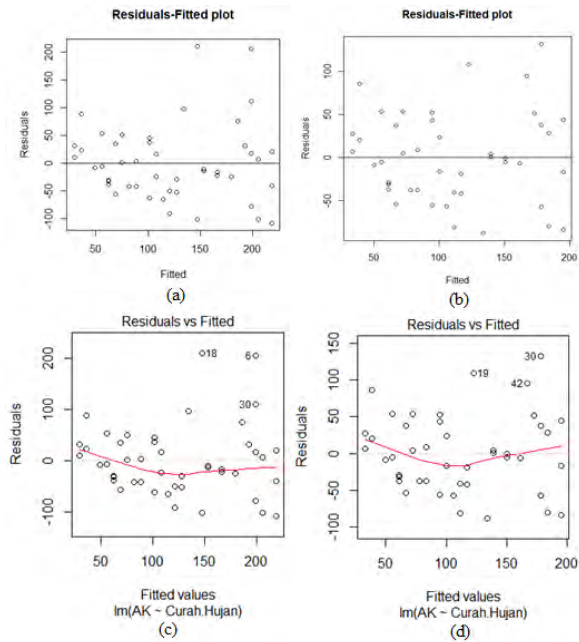
$$AK = -5,5695 * \text{Curah.Hujan} + 194,8720$$

dengan detail perbandingan antara persamaan lama dan baru (setelah eliminasi pencilan) terangkum dalam Tabel 5.13.

Tabel 5. 14 Perbandingan model H2a lama dan baru

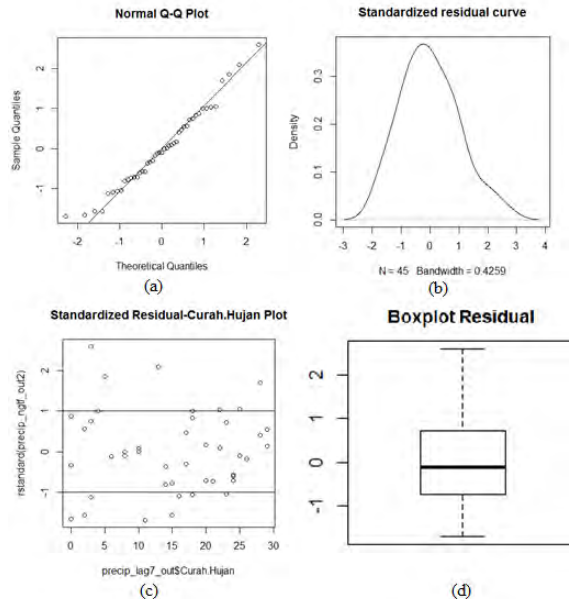
<i>Summary</i>	Korelasi negatif AK-Curah.Hujan lag 7	Sesudah eliminasi pencilan
Intercept	218,493	194,8720
Slope	-6,511	-5,5695
t value	-5,858	-6.423
p value	$5,08e^{-07}$	$8.85e^{-08}$
Signif. code	***0,001	***0,001
Residual std. error	68,52	52,29
DF	45	43
Multiple R-squared	0,4326	0,4897
Adjusted R-squared	0,42	0,4778
F-statistic	34,31	41,26

Nilai t-value, p-value, dan R^2 berubah semakin baik. Selanjutnya menguji kembali asumsi regresi.



Gambar 5. 20 Plot perbandingan heteroskedastisitas model H2a pasca eliminasi pencilan

Plot pada Gambar 5.20 (a) dan (c) merupakan kepunyaan model sebelum pengeliminasian pencilan. Setelah pencilan dibuang, nilai p-value meningkat menjadi 0,05536194. Sehingga varians residual dari model ini bisa dikatakan konstan.



Gambar 5. 21 Uji normalitas model H2a pasca eliminasi pencilan

Pada Gambar 5.21 (c) terdapat 14 dari 45 observasi yang berada diluar rentang nilai $[-1,1]$. Artinya, 68,89% observasi berada pada rentang nilai tersebut. Persentase tersebut telah melebihi 67,6%, sehingga dapat disimpulkan bahwa residual telah terdistribusi normal.

- (4) Evaluasi hasil peramalan pada data latih
Evaluasi hasil peramalan pada data latih menunjukkan nilai MAPE yang sangat besar, yakni 61,07%.
- (5) Melakukan peramalan dan evaluasi hasil peramalan
Evaluasi hasil peramalan pada data uji menunjukkan nilai MAPE yang sangat besar, yakni 79,38%.

5.4.2 Model H2b dengan variabel Ln(AK)

Berikut ini adalah tahap-tahap pembuatan model regresi untuk variabel Ln(AK) dan Curah.Hujan:

(1) Melakukan regresi

Persamaan regresi yang dihasilkan adalah:

$\text{Ln}(\text{AK}) = -0,05832 * \text{Curah.Hujan} + 5,39324$
dengan detail yang terangkum dalam Tabel 5.14.

Tabel 5. 15 Rangkuman model regresi H2b

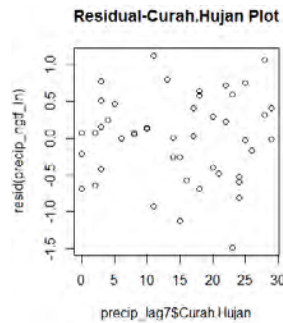
<i>Summary</i>	Korelasi negatif Ln(AK) - curah hujan lag 7
Intercept	5,393240
Slope	-0,058324
t value	-6,142
p value	$1,92e^{-07}$
Signif. code	***0,001
Residual std. error	0,5854
DF	45
Multiple R-squared	0,456
Adjusted R-squared	0,4439
F-statistic	37,72

(2) Interpretasi hasil

- t-tabel dengan degree of freedom 45 dan tingkat signifikansi 0,05 adalah 1.67943
- t-value persamaan regresi adalah -5,858. Karena $t\text{-value} < t\text{-tabel}$, maka H_0 ditolak
- p-value = $1,92e^{-07}$ menunjukkan pengaruh yang signifikan pada level 0,001.

(3) Uji asumsi regresi

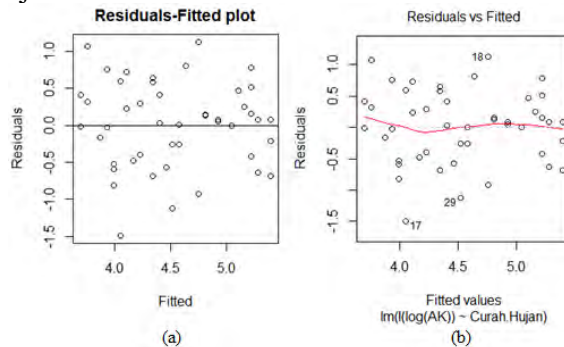
- Uji independensi residual



Gambar 5. 22 Plot residual model H2b

Plot residual pada Gambar 5.22 terdistribusi acak, tidak terlihat adanya sebuah fungsi yang menghubungkan residual dengan variabel Curah.Hujan.

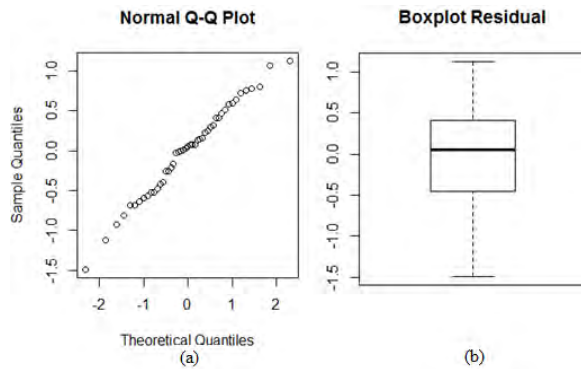
- Uji heteroskedastisitas



Gambar 5. 23 Plot untuk uji heteroskedastisitas model H2b

Pada Gambar 5.23, terlihat tidak ada kecenderungan data melebar atau menyempit. Nilai p-value dari tes statistik adalah 0,2578973 sehingga mengindikasikan varians telah konstan.

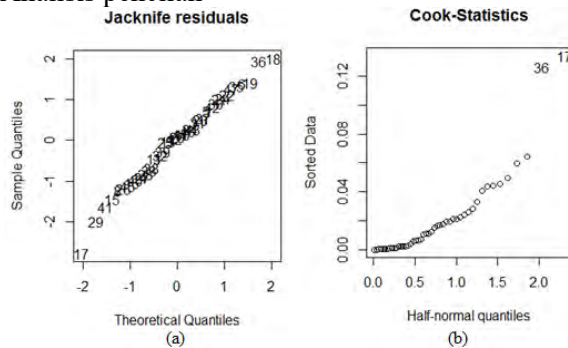
- Residual terdistribusi normal



Gambar 5. 24 Plot untuk uji normalitas residual model H2b

Residual terdistribusi normal karena berada di sekitar garis linear pada normal Q-Q plot (Gambar 5.24 (a)) dan memiliki rata-rata nol pada boxplot residual (Gambar 5.24 (b)).

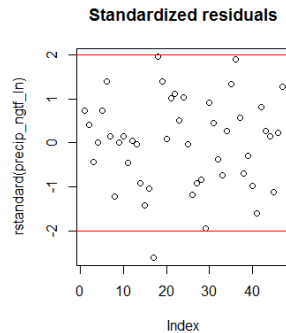
– Analisis pencilan



Gambar 5. 25 Plot untuk analisis pencilan model H2b

Plot Jackknife residual (Gambar 5.25 (a)) menampilkan index observasi pada plot Q-Q normal dalam bentuk angka. Grafik Cook-

Statistics (Gambar 5.25 (b)) menunjukkan observasi-observasi yang berpotensi menjadi pencilan dengan menganalisis nilai studentized residual-nya. Observasi yang mempunyai nilai studentized residual lebih dari 2,0 adalah observasi 36 dan 17.



Gambar 5. 26 Plot standardized residual model H2b

Standardized residual juga dapat digunakan untuk mendeteksi pencilan. Pencilan pada plot standardized residual adalah observasi yang berada di luar rentang nilai $[-2,2]$. Berbeda dengan hasil perhitungan cook statistic (Gambar 5.25 (b)), pada plot standardized residual (Gambar 5.26) hanya terdapat satu pencilan yang terdeteksi. Oleh karena itu, hanya observasi ke 17 yang akan diperlakukan sebagai pencilan.

- Analisis pasca eliminasi pencilan
Observasi yang menjadi pencilan ditunjukkan oleh Tabel 5.15.

Tabel 5. 16 Outlier H2b

Index Obs.	Ln(AK)	Curah.Hujan
17	2,564949	23

Persamaan regresi baru yang dihasilkan adalah:

$\text{Ln}(\text{AK}) = -0,055034 * \text{Curah.Hujan} + 5,377065$
 dengan detail perbandingan antara persamaan lama dan baru (setelah eliminasi pencilan) terangkum dalam Tabel 5.16.

Tabel 5. 17 Perbandingan model H2b lama dan baru

<i>Summary</i>	Korelasi negatif Ln(AK) - curah hujan lag 7	Sesudah eliminasi pencilan
Intercept	5,393240	5,377065
Slope	-0,058324	-0,055034
t value	-6,142	-6,158
p value	$1,92e^{-07}$	$1,98e^{-07}$
Signif. code	***0,001	***0,001
Residual std. error	0,5854	0,5461
DF	45	44
Multiple R- squared	0,456	0,4629
Adjusted R- squared	0,4439	0,4507
F-statistic	37,72	37,92

Pengeliminasian pencilan pada model tidak berpengaruh banyak terhadap perubahan nilai intercept dan slope. Sehingga pencilan tersebut dapat dibiarkan.

- (4) Evaluasi hasil peramalan pada data latih
 Evaluasi hasil peramalan pada data latih menunjukkan nilai MAPE 11,00%.

- (5) Melakukan peramalan dan evaluasi hasil peramalan
Evaluasi hasil peramalan pada data uji menunjukkan nilai MAPE 11,89%.

5.5 Pembuatan Model H3

5.5.1 Model H3a dengan variabel AK

Berikut ini adalah tahap-tahap pembuatan model regresi untuk variabel AK dan Kelembaban:

- (1) Melakukan regresi
Persamaan regresi yang dihasilkan adalah:
$$AK = -7,021 * \text{Kelembaban} + 700,500$$

dengan detail yang terangkum dalam Tabel 5.17.

Tabel 5. 18 Rangkuman model regresi H3a

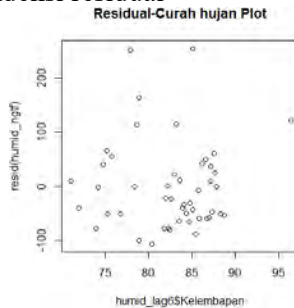
<i>Summary</i>	Korelasi negatif AK- kelembaban lag 6
Intercept	700,500
Slope	-7,021
t value	-3,030
p value	0,004
Signif. code	**0,01
Residual std. error	82,19
DF	46
Multiple R-squared	0,1664
Adjusted R-squared	0,1483
F-statistic	9,184

- (2) Interpretasi Hasil
- t-tabel dengan degree of freedom 46 dan tingkat signifikansi 0,05 adalah 1.67866

- t-value persamaan regresi adalah -3,030. Karena $t\text{-value} < t\text{-tabel}$, maka H_0 ditolak.
- p-value = 0,004 menunjukkan pengaruh yang signifikan pada level 0,01.

(3) Uji asumsi regresi

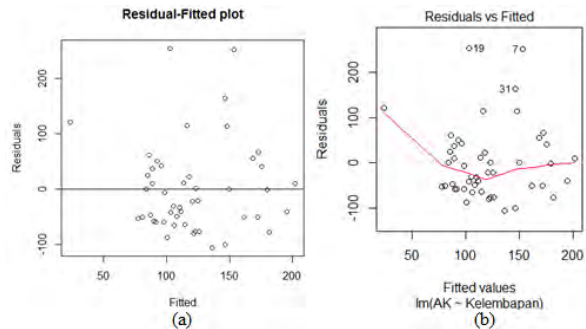
- Uji independensi residual



Gambar 5. 27 Plot residual model H3a

Plot residual pada Gambar 5.27 terdistribusi acak, tidak terlihat adanya sebuah fungsi yang menghubungkan residual dengan variabel Kelembaban. Selain itu korelasi antara residual dengan Kelembaban adalah $-2.453591e^{-17}$. Hal ini menunjukkan residual independen/bebas dari pengaruh variabel Kelembaban.

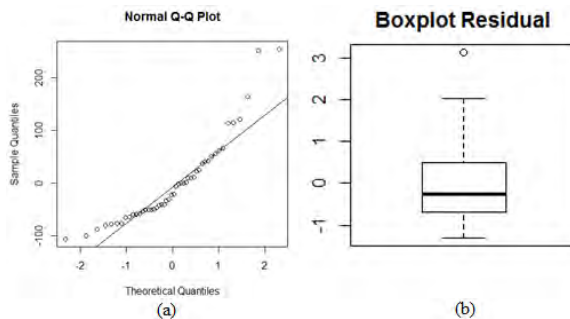
- Uji heteroskedastisitas



Gambar 5. 28 Plot untuk uji heteroskedastisitas model H3a

Pada Gambar 5.28, terlihat tidak ada kecenderungan data melebar atau menyempit. Nilai p-value dari tes statistik adalah 0,7186655 sehingga mengindikasikan varians yang telah konstan.

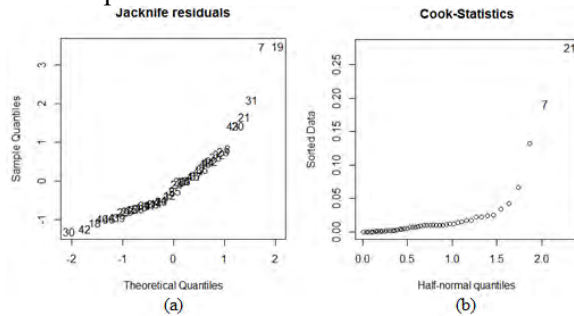
- Residual terdistribusi normal



Gambar 5. 29 Plot untuk uji normalitas residual model H3a

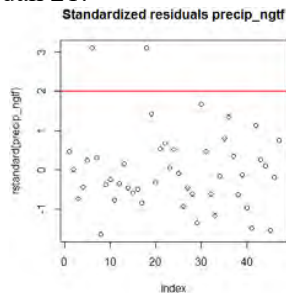
Residual mendekati distribusi normal karena berada di sekitar garis linear pada normal Q-Q plot (Gambar 5.29 (a)) dan memiliki rata-rata mendekati nol pada boxplot residual (Gambar 5.29 (b)).

– Analisis pencilan



Gambar 5. 30 Plot untuk analisis pencilan model H3a

Plot Jackknife residual (Gambar 5.30 (a)) menampilkan index observasi pada plot Q-Q normal dalam bentuk angka. Grafik Cook-Statistics (Gambar 5.30 (b)) menunjukkan observasi-observasi yang berpotensi menjadi pencilan dengan menganalisis nilai studentized residual-nya. Observasi yang mempunyai nilai studentized residual lebih dari 2,0 adalah observasi 7 dan 21.



Gambar 5. 31 Plot standardized residual model H3a

- Standardized residual juga dapat digunakan untuk mendeteksi pencilan. Pencilan pada plot standardized residual (Gambar 5.31) adalah observasi yang berada di luar rentang nilai $[-2,2]$.
- Analisis pasca eliminasi pencilan
Observasi yang menjadi pencilan ditunjukkan oleh Tabel 5.18.

Tabel 5. 19 Outlier H3a

Index Obs.	AK	Kelembaban
7	405	77,93548
21	145	96,40000

Persamaan regresi baru yang dihasilkan adalah:

$$AK = -7,590 * \text{Kelembaban} + 739,268$$

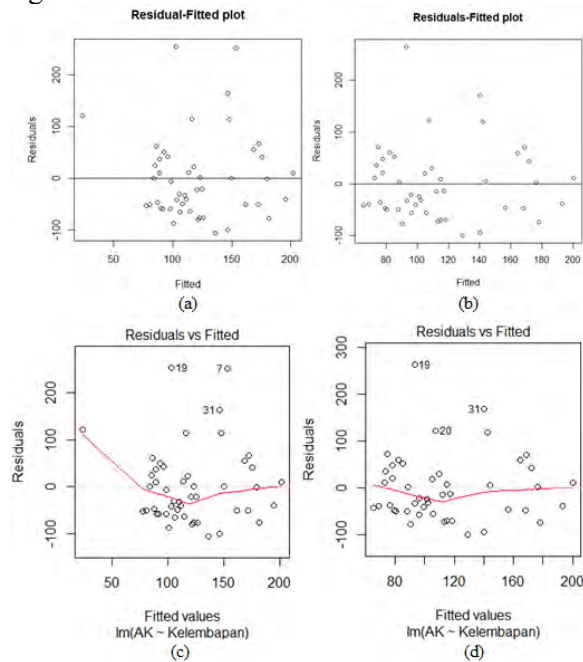
dengan detail perbandingan antara persamaan lama dan baru (setelah eliminasi pencilan) terangkum dalam Tabel 5.19.

Tabel 5. 20 Perbandingan model H3a lama dan baru

<i>Summary</i>	Korelasi negatif AK-Kelembaban lag 6	Sesudah eliminasi pencilan
Intercept	700,500	739.268
Slope	-7,021	-7,590
t value	-3,030	-3,396
p value	0,004	0,001458
Signif. code	**0,01	**0,01
Residual std. error	82,19	72,2
DF	46	44

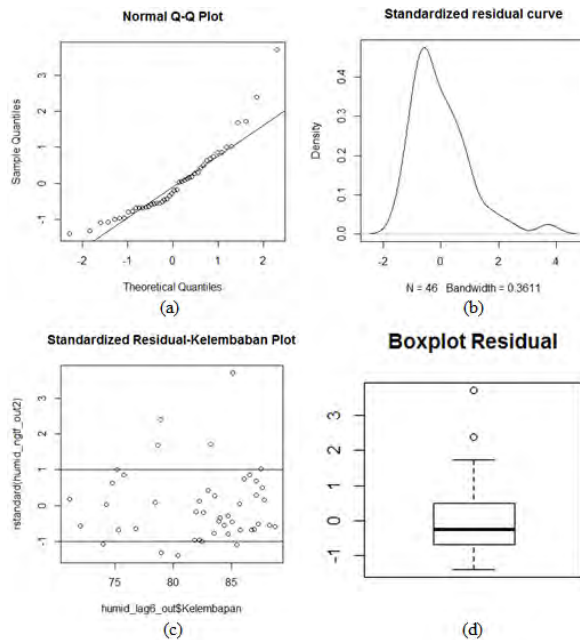
<i>Summary</i>		Korelasi negatif AK- Kelembaban lag 6	Sesudah eliminasi pencilan
Multiple R-squared	R	0,1664	0,2077
Adjusted R-squared	R	0,1483	0,1897
F-statistic		9,184	11,54

Nilai t-value, p-value, dan R^2 berubah semakin baik. Selanjutnya menguji kembali asumsi regresi.



Gambar 5. 32 Perbandingan heteroskedastisitas model H3a pasca eliminasi pencilan

Plot pada Gambar 5.32 (a) dan (c) merupakan kepunyaan model sebelum pengeliminasian pencilan. Setelah pencilan dibuang, nilai p-value meningkat menjadi 0,9658004 sehingga dapat dinyatakan bahwa varians tetap konstan.



Gambar 5. 33 Uji normalitas model H3a pasca eliminasi pencilan

Pada Gambar 5.33 (c) terdapat 9 dari 46 observasi yang berada diluar rentang nilai $[-1,1]$. Artinya, 80,43% observasi berada pada rentang nilai tersebut. Persentase tersebut telah melebihi 67,6%, sehingga dapat disimpulkan bahwa residual telah terdistribusi normal.

(4) Evaluasi hasil peramalan pada data latih

Evaluasi hasil peramalan pada data latih menunjukkan nilai MAPE yang sangat besar, yakni 80,33%.

- (5) Melakukan peramalan dan evaluasi hasil peramalan
Evaluasi hasil peramalan pada data uji menunjukkan nilai MAPE yang sangat besar, yakni 69,83%.

5.5.2 Model H3b dengan variabel Ln(AK)

Berikut ini adalah tahap-tahap pembuatan model regresi untuk variabel Ln(AK) dan Kelembaban:

- (1) Melakukan regresi

Persamaan regresi yang dihasilkan adalah:

$\text{Ln(AK)} = -0,06318 * \text{Kelembaban} + 9,740703$
dengan detail yang terangkum dalam Tabel 5.20.

Tabel 5. 21 Rangkuman model regresi H3b

<i>Summary</i>	Korelasi negatif AK- kelembaban lag 6
Intercept	9,740703
Slope	-0,06318
t value	-3,148
p value	0,00288
Signif. code	**0,01
Residual std. error	0,7121
DF	46
Multiple R-squared	0,1772
Adjusted R-squared	0,1594
F-statistic	9,91

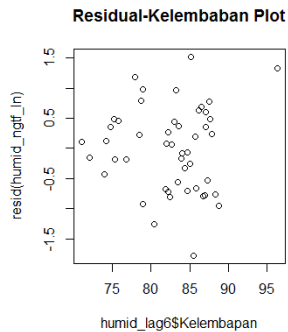
- (2) Interpretasi hasil

- t-tabel dengan degree of freedom 46 dan tingkat signifikansi 0,05 adalah 1.67866

- t-value persamaan regresi adalah -3,148. Karena $t\text{-value} < t\text{-tabel}$, maka H_0 ditolak.
- p-value = 0,00288 menunjukkan pengaruh yang signifikan pada level 0,01.

(3) Uji asumsi regresi

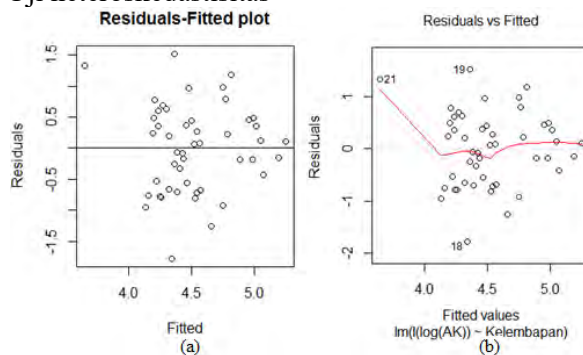
- Uji independensi residual



Gambar 5. 34 Plot residual model H3b

Plot terdistribusi acak, tidak terlihat adanya sebuah fungsi yang menghubungkan residual dengan variabel Kelembaban.

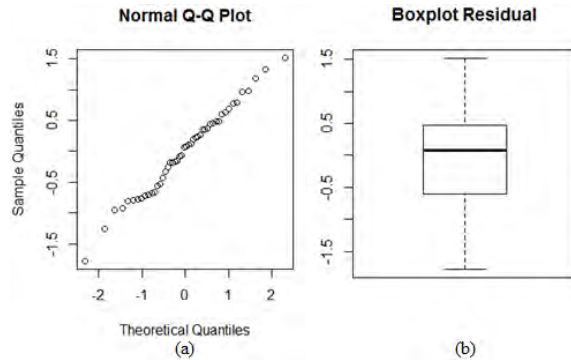
- Uji heteroskedastisitas



Gambar 5. 35 Plot untuk uji heteroskedastisitas model H3b

Pada Gambar 5.35, terlihat ada kecenderungan data sedikit menyempit. Namun nilai p-value dari tes statistik adalah 0,07700048 sehingga mengindikasikan varians telah konstan.

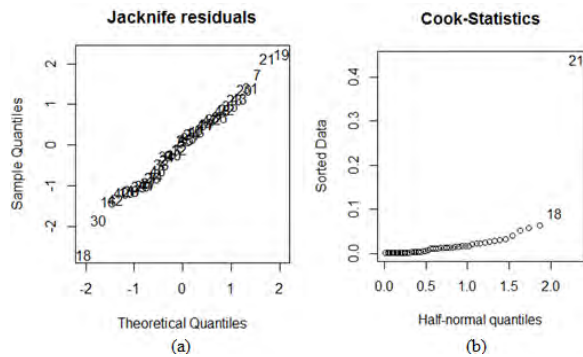
- Residual terdistribusi normal



Gambar 5. 36 Plot untuk uji normalitas residual model H3b

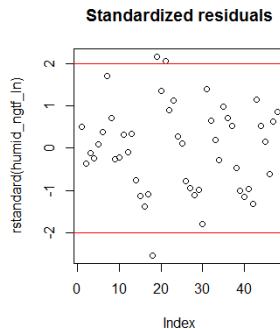
Residual terdistribusi normal karena membentuk pola garis linear pada normal Q-Q plot (Gambar 5.36 (a)) dan memiliki rata-rata nol pada boxplot residual (Gambar 5.36 (b)).

- Analisis pencilan



Gambar 5. 37 Plot untuk analisis pencilan model H3b

Plot Jackknife residual (Gambar 5.37 (a)) menampilkan index observasi pada plot Q-Q normal dalam bentuk angka. Grafik Cook-Statistics (Gambar 5.37 (b)) menunjukkan observasi-observasi yang berpotensi menjadi pencilan dengan menganalisis nilai studentized residual-nya. Observasi yang mempunyai nilai studentized residual lebih dari 2,0 adalah observasi 18 dan 21



Gambar 5. 38 Plot standardized residual model H3b

Standardized residual juga dapat digunakan untuk mendeteksi pencilan. Pencilan pada plot standardized residual adalah observasi yang berada di luar rentang nilai $[-2,2]$. Pencilan yang didapatkan dari plot standardized residual (Gambar 5.38) adalah sejumlah tiga. Namun berdasarkan cook-statistics (Gambar 5.37 (b)), yang berpotensi mempunyai pengaruh terhadap perubahan nilai slope dan intercept hanya observasi 18 dan 21.

- Analisis pasca eliminasi pencilan
Observasi yang menjadi pencilan ditunjukkan oleh Tabel 5.21.

Tabel 5. 22 Outlier H3b

Index Obs.	Ln(AK)	Kelembaban
18	2,564949	85,46667
21	4,976734	96,40000

Persamaan regresi baru yang dihasilkan adalah:

$\text{Ln(AK)} = -0,07558 * \text{Kelembaban} + 10,76881$
dengan detail perbandingan antara persamaan lama dan baru (setelah eliminasi pencilan) terangkum dalam Tabel 5.22.

Tabel 5. 23 Perbandingan model H3b lama dan baru

<i>Summary</i>	Korelasi negatif AK-kelembaban lag 6	Sesudah eliminasi pencilan
Intercept	9,740703	10,76881
Slope	-0,06318	-0,07558
t value	-3,148	-3,807

<i>Summary</i>	Korelasi negatif AK-kelembaban lag 6	Sesudah eliminasi pencilan
p value	0,00288	0,000432
Signif. code	**0,01	***0,001
Residual std. error	0,7121	0,6439
DF	46	44
Multiple R-squared	0,1772	0,2478
Adjusted R-squared	0,1594	0,2307
F-statistic	9,91	14,49

Pengeliminasian pencilan pada model tidak berpengaruh banyak terhadap perubahan nilai intercept dan slope. Sehingga pencilan tersebut dapat dibiarkan.

- (4) Evaluasi hasil peramalan pada data latih
Evaluasi hasil peramalan pada data latih menunjukkan nilai MAPE 13,67%.
- (5) Melakukan peramalan dan evaluasi hasil peramalan
Evaluasi hasil peramalan pada data uji menunjukkan nilai MAPE 8,88%.

5.6 Pembuatan Model H4

5.6.1 Model H4a dengan variabel AK

Berikut ini adalah tahap-tahap pembuatan model regresi untuk variabel AK dan tiga variabel cuaca (Temperatur, Curah.Hujan dan Kelembaban). Variabel cuaca yang digunakan pada model regresi linier berganda ini adalah

variabel-variabel yang juga digunakan pada tiga model regresi sederhana sebelumnya.

(1) Melakukan regresi

Persamaan regresi awal yang dihasilkan adalah:

$$AK = 789,649 - 36,744 * \text{Temperatur} - 6,274 * \text{Curah.Hujan} + 3,580 * \text{Kelembaban}$$

dengan detail yang terangkum dalam Tabel 5.23.

Tabel 5. 24 Rangkuman model regresi H4a

<i>Summary</i>	Temperat ur	Curah.Hu jan	Kelemba ban
Intercept	789,649		
Slope	-36,744	-6,274	3,580
Signifikansi	<0,05	<0,001	>0,1
t value	-2,127	-4,047	1,290
p value	0,039172	0,000212	0,203831
Residual std. error	66,4		
DF	43		
Multiple R- squared	0,4908		
Adjusted R- squared	0,4553		
F-statistic	13,82		

(2) Interpretasi Hasil dan Anova

Variabel temperatur dan curah hujan menunjukkan signifikansi berturut-turut 0,05 dan 0,001. Sedangkan variabel kelembaban tidak menunjukkan pengaruh yang signifikan, walaupun pada saat dilakukan regresi sederhana menunjukkan pengaruh yang signifikan.

Oleh karena temperatur, curah hujan dan kelembaban tidak memberikan pengaruh yang signifikan secara bersama-sama terhadap variabel AK, maka H_0 diterima.

Di dalam R, *partial F-test* dapat dilakukan dengan cara membuat model penuh dan model yang dikurangi variabelnya (model tereduksi) secara terpisah, lalu membandingkannya menggunakan ANOVA. Dalam hal ini, variabel kelembaban akan dikeluarkan dari model ($H_0: \beta_3=0$). Tabel 5.24 adalah ringkasan dari model yang tidak melibatkan variabel kelembaban.

Tabel 5. 25 Ringkuman model regresi H4a tereduksi

Summary	Temperatur	Curah.Hujan
Intercept	894,167	
Slope	-29,374	-5,190
Signifikansi	<0,1	<0,001
t value	-1,789	-3,954
p value	0,080579	0,000275
Residual std. error	66,9	
DF	44	
Multiple R-squared	0,4711	
Adjusted R-squared	0,447	
F-statistic	19,59	

Variabel curah hujan masih mempunyai level signifikansi yang sama, sementara variabel temperatur mengalami penurunan level signifikansi menjadi

hanya 0,1. Anova antara model dengan variabel cuaca penuh dengan model tereduksi menunjukkan nilai $F=1,665$ dan $p\text{-value}=0,2038$, sehingga hipotesis nol tidak dapat ditolak pada level signifikansi 0,05. Jadi, variabel kelembaban tidak memberikan kontribusi informasi yang signifikan terhadap AK ketika variabel temperatur dan curah hujan sudah masuk di dalam model.

Karena variabel temperatur tidak mempunyai level signifikan $<0,05$ pada model tereduksi tersebut, maka dilakukan analisis ANOVA kedua. Analisis ini membandingkan model yang hanya terdiri dari variabel Curah.Hujan sebagai prediktor dengan model yang melibatkan Temperatur dan Curah.Hujan. Hasil perhitungan menunjukkan nilai $F=3,1988$ dan $p\text{-value} = 0,08058$, sehingga H_0 tidak dapat ditolak pada level signifikansi 0,05. Jadi, variabel Temperatur juga tidak memberikan kontribusi informasi yang signifikan terhadap AK ketika variabel Curah.Hujan sudah masuk di dalam model.

Sebagai perbandingan, ketika variabel Curah.Hujan dikeluarkan dari model, nilai $F=15,634$ dan $p\text{-value}=0,0002755$. Sehingga H_0 ($\beta_2 = 0$) ditolak. Variabel Curah.Hujan cukup memberikan kontribusi informasi yang signifikan pada level 0,05 terhadap variabel AK walaupun sudah ada variabel Temperatur.

(3) Uji asumsi regresi

Karena model yang dihasilkan dari percobaan membuat model regresi berganda sama dengan model regresi sederhana yang hanya melibatkan

variabel Curah.Hujan sebagai prediktor, maka hasil pengujian yang dilakukan sama dengan model H2a.

- (4) Evaluasi hasil peramalan pada data latih
Evaluasi hasil peramalan pada data latih menunjukkan nilai MAPE 61,07%.
- (5) Melakukan peramalan dan evaluasi hasil peramalan
Evaluasi hasil peramalan pada data uji menunjukkan nilai MAPE 79,38%.

5.6.2 Model H4b dengan variabel Ln(AK)

Berikut ini adalah tahap-tahap pembuatan model regresi untuk variabel Ln(AK) dan tiga variabel cuaca (Temperatur, Curah.Hujan dan Kelembaban). Variabel cuaca yang digunakan pada model regresi linier berganda ini adalah variabel-variabel yang juga digunakan pada tiga model regresi sederhana sebelumnya

- (1) Melakukan regresi
Persamaan regresi awal yang dihasilkan adalah:

$$\text{Ln(AK)} = -0,24261 * \text{Temperatur} - 0,05729 * \text{Curah.Hujan} + 0,02496 * \text{Kelembaban} + 9,06301$$
 dengan detail yang terangkum dalam Tabel 5.25.

Tabel 5. 26 Rangkuman model regresi H4b

Summary	Temperatur	Curah.Hujan	Kelembaban
Intercept	9,06301		
Slope	-0,24261	-0,05729	0,02496
Signifikansi	>0,1	<0,001	>0,1
t value	-1,609	-4,232	1,030
p value	0,115020	0,000119	0,308558

<i>Summary</i>	Temperat ur	Curah.Hu jan	Kelemba ban
Residual std. error	0,5798		
DF	43		
Multiple R- squared	0,49		
Adjusted R- squared	0,4544		
F-statistic	13,77		

(2) Interpretasi hasil dan Anova

Menggunakan dataset yang sama seperti semua model regresi sederhana, didapatkan hasil yang berbeda. Hanya variabel Curah.Hujan pada lag 7 yang menunjukkan signifikansi pada level 0,001. Variabel temperatur dan kelembaban tidak menunjukkan pengaruh yang signifikan, walaupun pada saat dilakukan regresi sederhana keduanya menunjukkan pengaruh yang signifikan. Oleh karena temperatur, curah hujan dan kelembaban tidak memberikan pengaruh yang signifikan secara bersama-sama terhadap variabel $\ln(AK)$, maka H_0 diterima.

Di dalam R, partial F-test dapat dilakukan dengan cara membuat model penuh dan model yang dikurangi variabelnya (model tereduksi) secara terpisah, lalu membandingkannya menggunakan ANOVA. Variabel kelembaban akan coba dikeluarkan terlebih dahulu karena memiliki nilai p-value yang paling tidak signifikan.

Tabel 5. 27 Rangkuman model regresi H4b tereduksi

<i>Summary</i>	Temperatur	Curah.Hujan
Intercept	9,79184	
Slope	-0,19122	-0,04973
Signifikansi	>0,1	<0,001
t value	-1,342	-4,368
p value	0,18632	7,52e ⁻⁰⁵
Residual std. error	0,5802	
DF	44	
Multiple R-squared	0,4774	
Adjusted R-squared	0,4537	
F-statistic	20,1	

Variabel curah hujan masih mempunyai level signifikansi yang sama namun dengan p-value yang meningkat, sementara variabel temperatur menjadi semakin tidak signifikan (mengalami penurunan level signifikansi). Anova antara model dengan variabel cuaca penuh dengan model tereduksi menunjukkan nilai $F=1,0618$ ($p\text{-value} = 0,3086$), sehingga hipotesis nol tidak dapat ditolak pada level signifikansi 0,05. Jadi, variabel kelembaban tidak memberikan kontribusi informasi yang signifikan terhadap $\ln(AK)$ ketika variabel Temperatur dan Curah.Hujan sudah masuk di dalam model.

Karena variabel temperatur tidak mempunyai level signifikan <0,05 pada model tereduksi tersebut, maka dilakukan ANOVA kedua. Analisis ini

membandingkan model yang hanya terdiri dari variabel Curah.Hujan sebagai prediktor dengan model yang melibatkan Temperatur dan Curah.Hujan. Hasil perhitungan menunjukkan nilai $F=1,8023$ ($p\text{-value} = 0,1863$), sehingga H_0 tidak dapat ditolak pada level signifikansi 0,05. Jadi, variabel temperatur tidak memberikan kontribusi informasi yang signifikan terhadap $\ln(AK)$ ketika variabel Curah.Hujan sudah masuk di dalam model.

(3) Uji asumsi regresi

Karena model yang dihasilkan dari percobaan membuat model regresi berganda sama dengan model regresi sederhana yang hanya melibatkan variabel Curah.Hujan sebagai prediktor, maka hasil pengujian yang dilakukan sama dengan model H2b.

(4) Evaluasi hasil peramalan pada data latih

Evaluasi hasil peramalan pada data latih menunjukkan nilai MAPE 11,00%.

(5) Melakukan peramalan dan evaluasi hasil peramalan

Evaluasi hasil peramalan pada data uji menunjukkan nilai MAPE 11,89%.

Halaman ini sengaja dikosongkan.

BAB VI

HASIL DAN PEMBAHASAN

Bab ini berisi hasil percobaan (implementasi) dan pembahasan mengenai hasil tersebut, antara lain analisis dan evaluasi hasil regresi.

6.1 Data Hasil

Data hasil merupakan data hasil percobaan pada setiap model, dikelompokkan dan dijelaskan pada sub bab berikutnya berdasarkan variabel cuaca yang digunakan sebagai variabel dependen model.

6.1.1. Model H1

Perbandingan antara model H1a dan H1b dapat dilihat pada Tabel 6.1. Penjelasan untuk kolom H1a terdapat pada sub bab 5.3.1, sedangkan kolom H1b terdapat pada sub bab 5.3.2.

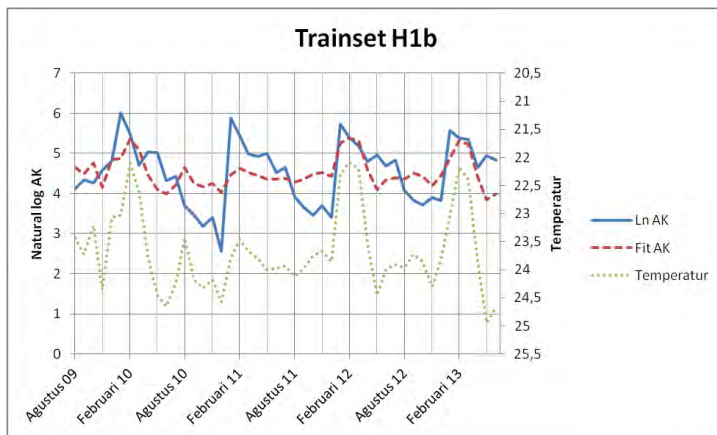
Tabel 6. 1 Perbandingan model H1a dan H1b

	H1a	H1b
Persamaan	$AK = -60,22 * \text{Temperatur} + 1536,15$	$\text{Ln}(AK) = -0,5412 * \text{Temperatur} + 17,33505$
Uji t	H_0 ditolak	H_0 ditolak
p-value	<.001	<.001
R^2	0,3606	0,2342
Independensi residual	ya	ya
Heteroskedastisitas	tidak	tidak

	H1a	H1b
Normalitas residual	ya	ya
MAPE data latih	75,41%	13,60%
MAPE data uji	61,22%	9,02%

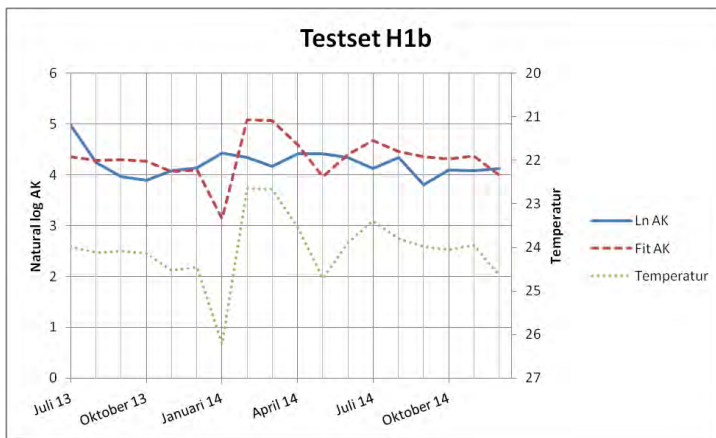
Nilai koefisien determinasi (R^2) model H1a lebih baik daripada H1b karena telah mengalami pengeliminasian pencilan. Pengeliminasian pencilan dapat meningkatkan kemampuan variabel independen dalam menjelaskan variabel dependen.

Nilai MAPE dari data latih dan data uji milik model H1a tergolong sangat tinggi, yakni berturut-turut 75,41% dan 61,22%. Sedang model yang menggunakan variabel natural log AK (H1b) memiliki nilai MAPE yang baik dan sangat baik, yakni 13,60% dan 9,02%. Visualisasi grafik dari hasil prediksi model H1b menggunakan data latih terdapat pada Gambar 6.1.



Gambar 6.1 Hasil prediksi model H1b menggunakan data latih

Pada Gambar 6.1 terlihat bahwa temperatur pada tahun 2011 tidak sefluktuatif temperatur pada tahun-tahun lainnya. Padahal angka kejadian pada tahun selain 2011 cukup mengikuti pola temperatur. Metode regresi OLS yang ide pokoknya adalah meminimalkan SSE, menyebabkan hasil prediksi cenderung memiliki varian yang lebih kecil dibandingkan dengan data aktual.



Gambar 6.2 Hasil prediksi model H1b menggunakan data uji

Hasil prediksi menggunakan data uji, yakni pada Gambar 6.2, menunjukkan pola variabel AK yang tidak mengikuti pola temperatur. Sehingga hasil prediksi berada jauh dari nilai aktualnya di beberapa titik.

Berdasarkan gambar 6.1 dan 6.2, secara kasat mata hasil prediksi menggunakan OLS tidak terlalu pas dengan data aktualnya. Namun nilai MAPE pada model H1b lebih kecil daripada MAPE pada model H1a. Sehingga diantara kedua model tersebut, model H1b lebih baik dari model H1a.

6.1.2. Model H2

Perbandingan antara model H1a dan H1b dapat dilihat pada Tabel 6.2. Penjelasan untuk kolom H2a terdapat pada sub bab 5.4.1, sedangkan kolom H2b terdapat pada sub bab 5.4.2

Tabel 6. 2 Perbandingan model H2a dan H2b

	H2a	H2b
Persamaan	$AK = -5,5695$ $*Curah.Hujan +$ $194,8720$	$Ln(AK) = -0,05832$ $*Curah.Hujan +$ $5,39324$
Uji t	H_0 ditolak	H_0 ditolak
p-value	<.001	<.001
R^2	0,4778	0,4439
Independensi residual	ya	ya
Heteroskedastisitas	tidak	tidak
Normalitas residual	ya	ya
MAPE data latih	61,07%	11,00%
MAPE data uji	79,38%	11,89%

Nilai koefisien determinasi (R^2) model H2a lebih baik daripada H2b karena telah mengalami pengeliminasian pencilan. Pengeliminasian pencilan dapat meningkatkan kemampuan variabel independen dalam menjelaskan variabel dependen.

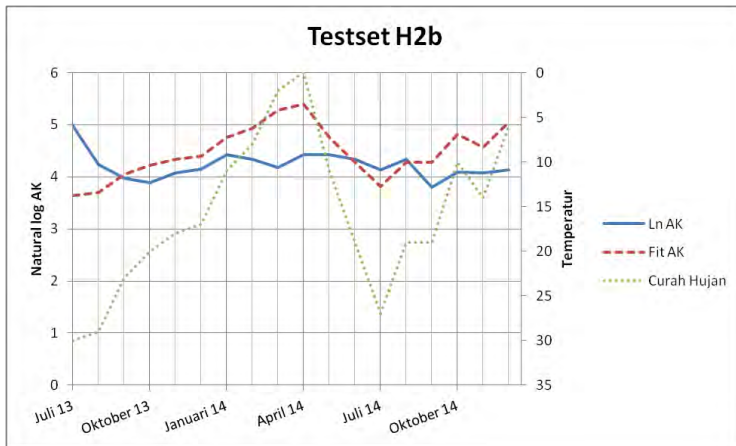
Nilai MAPE dari data latih dan data uji milik model H2a tergolong sangat tinggi, yakni berturut-turut 61,07% dan 79,38%. Sedang model yang menggunakan variabel natural log AK (H2b) memiliki nilai MAPE yang baik, yakni 11,00%

dan 11,89%. Visualisasi grafik dari hasil prediksi model H1b menggunakan data latih terdapat pada Gambar 6.3.



Gambar 6.3 Hasil prediksi model H2b menggunakan data latih

Pada Gambar 6.3 terlihat bahwa curah hujan pada tahun 2011 cenderung lebih rendah daripada curah hujan pada tahun-tahun lainnya. Padahal angka kejadian pada tahun 2011 tidak lebih rendah daripada tiga tahun lainnya. Dibanding dengan model H1b dan H3b, hasil prediksi H2b lebih baik. Ia dapat mendekati nilai tertinggi dan terendah variabel AK lebih baik dan karenanya memiliki nilai MAPE terkecil untuk data latih.



Gambar 6.4 Hasil prediksi model H2b menggunakan data uji

Hasil prediksi menggunakan data uji, yakni pada Gambar 6.4, menunjukkan pola variabel AK yang tidak menuruti pola curah hujan. Sehingga hasil prediksi berada jauh dari nilai aktualnya di beberapa titik.

Berdasarkan gambar 6.3 dan 6.4, periode puncak dari kedua variabel tidak selalu sama. Sehingga hasil prediksi menggunakan OLS tidak terlalu pas dengan data aktualnya. Namun, sama seperti pada model H1, nilai MAPE pada model H2b lebih kecil daripada MAPE pada model H2a. Sehingga diantara kedua model tersebut, model H2b lebih baik dari model H2a

6.1.3. Model H3

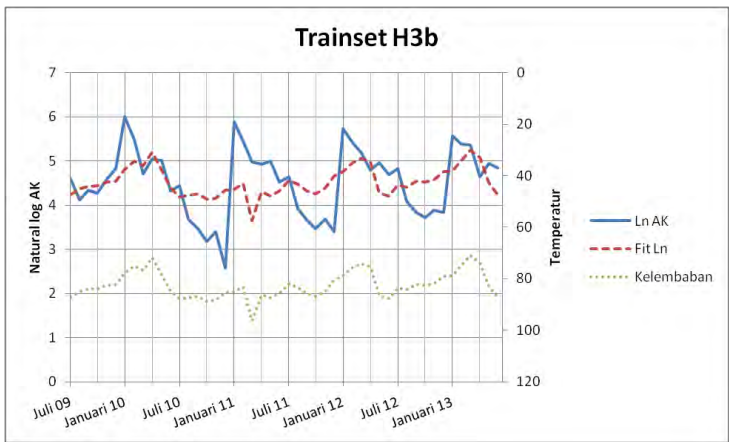
Perbandingan antara model H1a dan H1b dapat dilihat pada Tabel 6.3. Penjelasan untuk kolom H3a terdapat pada sub bab 5.5.1, sedangkan kolom H3b terdapat pada sub bab 5.5.2

Tabel 6. 3 Perbandingan model H3a dan H3b

	H3a	H3b
Persamaan	$AK = -7,590$ $*Kelembaban + 739,268$	$\ln(AK) = -0,06318$ $*Kelembaban + 9,740703$
Uji t	H_0 ditolak	H_0 ditolak
p-value	<.01	<.01
R^2	0,1897	0,1594
Independensi residual	ya	ya
Heteroskedastisitas	tidak	tidak
Normalitas residual	ya	ya
MAPE data latih	80,33%	13,67%
MAPE data uji	69,83%	8,88%

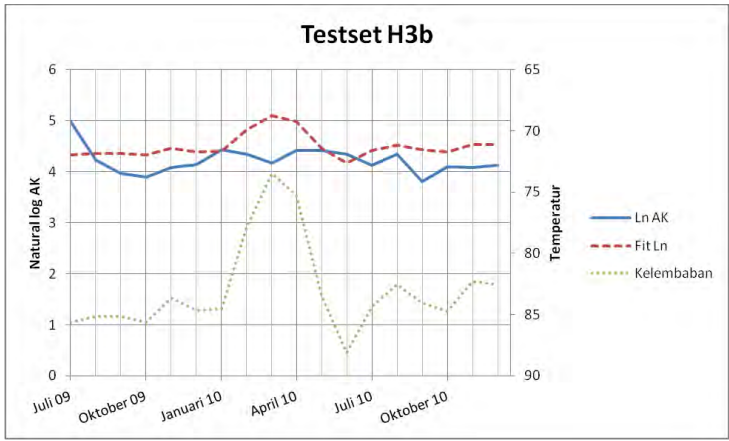
Nilai koefisien determinasi (R^2) model H3a lebih baik daripada H3b karena telah mengalami pengeliminasian pencilan. Pengeliminasian pencilan dapat meningkatkan kemampuan variabel independen dalam menjelaskan variabel dependen.

Nilai MAPE dari data latih dan data uji milik model H3a tergolong sangat tinggi, yakni berturut-turut 80,33% dan 69,83%. Sedang model yang menggunakan variabel natural log AK (H3b) memiliki nilai MAPE yang baik, yakni 13,67% dan 8,88%. Visualisasi grafik dari hasil prediksi model H3b menggunakan data latih terdapat pada Gambar 6.5.



Gambar 6.5 Hasil prediksi model H3b menggunakan data latih

Pada Gambar 6.5 terlihat bahwa rata-rata kelembaban pada tahun 2011 lebih rendah daripada kelembaban pada tahun-tahun lainnya. Padahal angka kejadian pada tahun selain 2011 cukup mengikuti pola kelembaban.



Gambar 6.6 Hasil prediksi model H3b menggunakan data uji

Hasil prediksi menggunakan data uji, yakni pada Gambar 6.6, menunjukkan pola variabel AK yang tidak mengikuti pola kelembaban. Sehingga hasil prediksi berada jauh dari nilai aktualnya di beberap titik.

Berdasarkan Gambar 6.5 dan 6.6, secara kasat mata hasil prediksi menggunakan OLS tidak terlalu pas dengan data aktualnya. Namun nilai MAPE pada model H3b lebih kecil daripada MAPE pada model H3a. Sehingga diantara kedua model tersebut, model H3b lebih baik dari model H3a.

6.2 Pembahasan/Diskusi

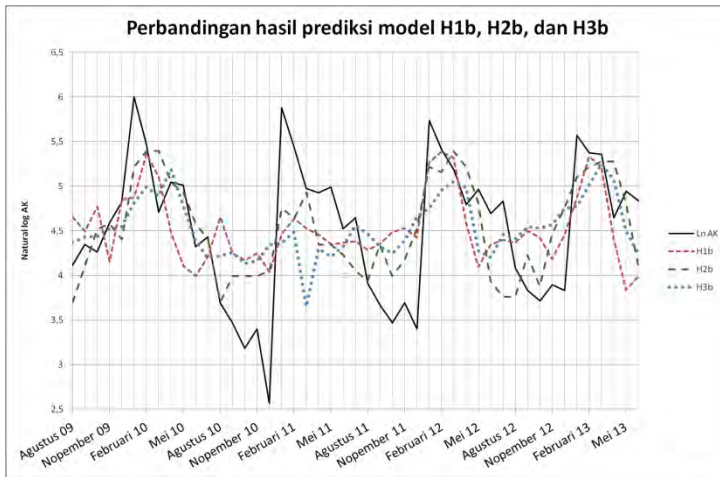
Koefisien determinasi dari model H1a, H2a, dan H3a (kelompok model A) masing-masing lebih baik daripada model H1b, H2b, dan H3b (kelompok model B). Hal ini karena koefisien determinasi dari kelompok model A mengalami peningkatan setelah dilakukan proses analisis pencilan. Koefisien determinasi yang paling baik adalah milik model dengan variabel independen berupa curah hujan, yakni 0,4778 untuk model H2a dan 0,4439 untuk model H2b.

Sedangkan untuk asumsi regresi, kesemuanya memenuhi asumsi regresi OLS. Walaupun pada saat melakukan percobaan, hasil awal model H2a terindikasi menyalahi asumsi homoskedastisitas. Secara teori heteroskedastisitas tidak dapat dibenahi dengan eliminasi pencilan, melainkan dengan melakukan transformasi variabel. Namun setelah dilakukan analisis pencilan, terjadi perbaikan nilai p-value. Sehingga model H2a tidak lagi menyalahi asumsi tersebut.

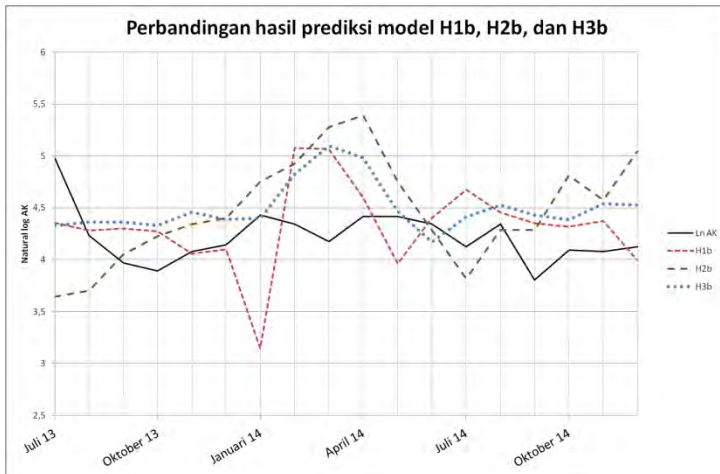
Satu-satunya yang menjadi permasalahan pada masing-masing model adalah nilai MAPE kelompok model A yang jauh lebih tinggi daripada nilai MAPE kelompok model B. Hal ini dikarenakan kelemahan dari pengukuran MAPE, yakni

bias terhadap hasil prediksi yang lebih rendah dari nilai aktual dan hanya relevan terhadap data berskala rasio [21]. Selain itu, varians dari variabel AK jauh lebih tinggi daripada varians variabel cuaca. Hal ini karena cuaca di kabupaten Malang cenderung stabil, hanya berada pada rentang nilai yang sempit. Variabel cuaca yang rentang nilainya paling besar adalah variabel curah hujan. Sehingga kemampuan curah hujan dalam menjelaskan variabilitas AK merupakan yang paling baik daripada dua variabel cuaca yang lain. Ini sesuai dengan pernyataan sebelumnya dimana koefisien determinasi model H2 adalah yang tertinggi.

Terkait permasalahan mengenai varians data ini, terdapat studi mengenai regresi lainnya yang pernah dilakukan dengan mengubah variabel dependen menjadi bentuk natural log untuk menstabilkan varians [28] [29]. Hal inilah yang mendasari variabel dependen pada kelompok model B menggunakan bentuk natural log AK. Gambar 6.7 dan Gambar 6.8 merupakan plot hasil prediksi model H1b, H2b, dan H3b dengan data aktual, berturut-turut dengan menggunakan data latih dan data uji.



Gambar 6.7 Perbandingan hasil prediksi model H1b, H2b, dan H3b menggunakan data latih



Gambar 6.8 Perbandingan hasil prediksi model H1b, H2b, dan H3b menggunakan data uji

Model regresi kelompok A yang melibatkan variabel curah hujan sebagai variabel independen (H2a) memiliki nilai MAPE paling kecil untuk data latih, yakni 61,07%. Gambar 6.7 menunjukkan hasil prediksi model H2b juga merupakan yang paling mendekati nilai aktual dengan nilai MAPE sebesar 11,00%. Namun demikian, model H2a dan H2b juga merupakan pemilik nilai MAPE terbesar untuk data uji, yakni berturut-turut 79,38% dan 11,89%. Hal ini karena pola data AK aktual pada data uji tidak sefluktuatif AK pada data latih. Sehingga untuk model H2 yang mempunyai tingkat fluktuatif paling tinggi pada data latih, berbalik mempunyai performa buruk pada data uji.

Perbedaan pola data uji dengan data latih mungkin berhubungan dengan hasil studi lain yang menyatakan bahwa agaknya kasus demam berdarah di Indonesia mempunyai siklus tahunan dan siklus lima tahunan [10]. Oleh karena itu, ketersediaan data yang lebih banyak akan sangat membantu dalam menemukan faktor-faktor lingkungan optimal yang mendukung terjadinya wabah demam berdarah. Faktor-faktor tersebut tidak hanya berguna dalam menghasilkan prediksi yang lebih akurat, namun juga dapat digunakan sebagai acuan untuk menentukan masa intervensi atau upaya pencegahan dini wabah demam berdarah.

6.3 Usulan Pemilihan/Penggunaan Metode Regresi

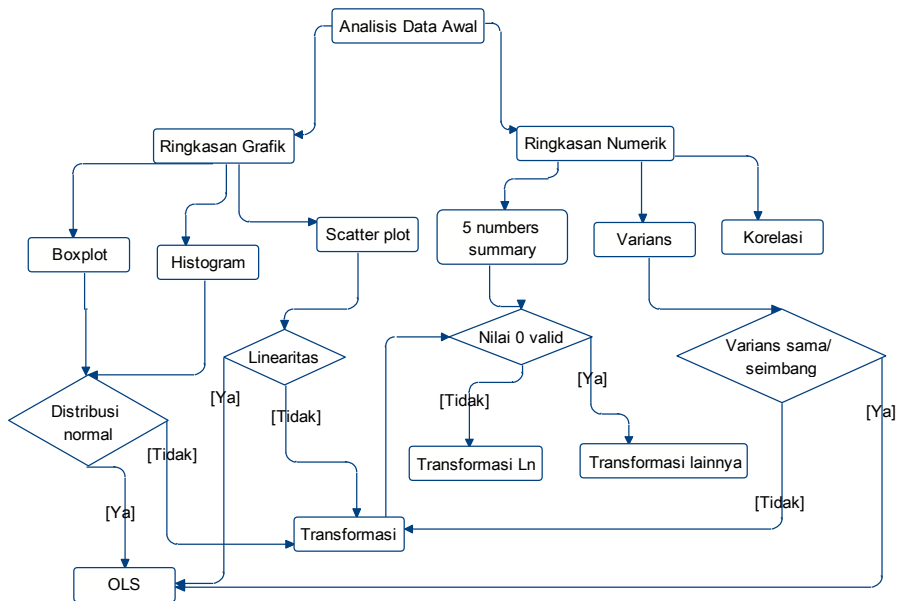
Penggunaan metode Regresi Ordinary Least Square tidak lepas dari beberapa asumsi, sebagai berikut:

1. Variabel independen/eksplanatoris/regressor dipastikan tidak terdapat kesalahan pada pengukurannya. Sehingga tingkat kesalahan terbebaskan seluruhnya pada variabel dependen [28].

2. Baik untuk data yang mempunyai varians variabel respon yang sama atau tidak berbeda jauh dengan varians variabel-variabel independennya. Hal ini penting karena variabilitas regressor harus bisa menjelaskan variabilitas variabel respon. Ketika varians antara keduanya timpang, maka hasil peramalan cenderung tidak dapat menyentuh titik maksimum dan minimum dari variabel respon.

Transformasi natural log merupakan salah satu bentuk transformasi yang paling umum digunakan. Namun penggunaannya juga dibatasi oleh beberapa kondisi berikut:

1. Transformasi natural log pada variabel respon mengasumsikan bahwa faktor kesalahan masuk ke dalam model secara multiplikatif. Artinya, setiap kenaikan sebanyak 1 unit pada variabel x_1 mengakibatkan peningkatan y sebanyak e^{β_1} (apabila model regresinya adalah $\log \hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1$).
2. Transformasi dilakukan pada variabel respon jika varians variabel respon jauh lebih besar dari varians regressor. Apabila justru varians regressor yang tidak seimbang dengan yang lain, maka yang harus ditransformasi adalah regressor yang bersangkutan.
3. Transformasi logaritmik tidak dapat dilakukan pada data yang mempunyai nilai nol mutlak.



Gambar 6. 9 Bagan keputusan menggunakan model OLS atau transformasi variabel

Pemilihan regresi OLS atau regresi dengan transformasi dapat ditentukan pada fase analisis data awal. Ringkasan grafik dan ringkasan numerik dari data dilakukan untuk mendapatkan petunjuk pemilihan model regresi yang tepat.

Ringkasan grafik pada data yang sering dilakukan adalah boxplot, histogram, dan scatter plot. Boxplot dan histogram berfungsi untuk melihat sebaran data, yakni apakah data sudah terdistribusi normal. Sedangkan scatter plot digunakan untuk melihat seberapa besar kemungkinan kedua variabel mempunyai hubungan linier yang kuat melalui “bentuk” dari sebaran data.

Ringkasan numerik pada data yang sering dilakukan adalah 5 *numbers summary*, terdiri dari Minimum, Quartil 1 (Q1),

Median, Mean, Quartil 3 (Q3) dan Maksimum, varians dan korelasi. Informasi yang didapatkan dari ringkasan numerik ini antara lain adanya nilai nol yang valid ataukah merupakan *missing value*, perbandingan varians antar variabel, dan angka korelasi kedua variabel.

OLS baik untuk dilakukan apabila semua variabel telah terdistribusi normal dan mempunyai kemungkinan yang besar untuk berhubungan linier, serta varians dari variabel yang terlibat dalam model sama, seimbang atau tidak terlalu berbeda jauh.

Sedangkan model regresi dengan transformasi variabel cocok digunakan apabila ditemukan variabel yang tidak berdistribusi normal, cenderung tidak memiliki hubungan linear yang dianggap cukup kuat dan varians yang tidak seimbang/terlalu berbeda jauh dengan variabel lainnya. Bentuk transformasi variabel ada banyak macamnya. Namun jika hendak menggunakan transformasi natural log, harus dipastikan bahwa variabel yang akan ditransformasi tidak mempunyai nilai nol.

Halaman ini sengaja dikosongkan.

LAMPIRAN A

Data mentah angka kejadian demam berdarah di kabupaten malang:

- Data asli

Bulan	Tahun					
	2009	2010	2011	2012	2013	2014
Januari	95	405	68	21	262	84
Februari	109	239	22	22	216	77
Maret	85	111	29	12	212	65
April	71	155	14	21	104	83
Mei	112	150	13	8	140	83
Juni	122	75	0	0	126	77
Juli	99	84	40	0	145	62
Agustus	61	40	8	3	69	77
September	77	32	14	0	53	45
Oktober	71	24	0	2	49	60
November	98	30	0	3	59	59
Desember	124	13	0	7	63	62
TOTAL	1124	1358	208	99	1498	834

- Data hasil pra-proses menggunakan interpolasi

Bulan	Tahun					
	2009	2010	2011	2012	2013	2014
Januari	95	405	357	310	262	84
Februari	109	239	231	224	216	77
Maret	85	111	145	178	212	65
April	71	155	138	121	104	83
Mei	112	150	147	143	140	83
Juni	122	75	92	109	126	77
Juli	99	84	104	125	145	62
Agustus	61	40	50	59	69	77
September	77	32	39	46	53	45
Oktober	71	24	32	41	49	60
November	98	30	40	49	59	59
Desember	124	13	30	46	63	62
TOTAL	1124	1358	1405	1451	1498	834

LAMPIRAN B

Periode	AK	Temperature	Curah Hujan	Kelembapan
Jan-09	95	23,43226	29	87,12903
Feb-09	109	23,72857	22	85,07143
Mar-09	85	23,21935	15	84,06452
Apr-09	71	24,35667	14	83,93333
Mei-09	112	23,07097	17	82,6129
Jun-09	122	23,02333	3	82,23333
Jul-09	99	22,11613	0	77,93548
Agust-09	61	22,60968	0	75,16129
Sep-09	77	23,76333	6	76,76667
Okt-09	71	24,45161	8	71,96774
Nop-09	98	24,65667	14	78,43333
Des-09	124	24,24839	17	84,74194
Jan-10	405	23,43226	29	87,82258
Feb-10	239	24,18214	24	87,35714
Mar-10	111	24,32581	24	86,77419
Apr-10	155	24,18333	24	88,76667
Mei-10	150	24,57742	23	88,32258
Jun-10	75	23,78333	11	85,46667
Jul-10	84	23,46452	13	85,12903
Agust-10	40	23,67419	8	83,22581
Sep-10	32	23,79667	18	96,4
Okt-10	24	23,99677	18	86,14516
Nop-10	30	23,9675	20	87,53333

B-2

Des-10	13	23,9371	23	85,72581
Jan-11	357	24,11129	25	81,96774
Feb-11	231	23,99107	18	83,49107
Mar-11	145	23,75	24	85,79839
Apr-11	138	23,66583	21	86,95833
Mei-11	147	23,85645	15	84,71774
Jun-11	92	22,30667	3	80,39167
Jul-11	104	22,08145	4	78,91935
Agust-11	50	22,20806	0	75,74194
Sep-11	39	23,54417	3	74,2
Okt-11	32	24,4629	10	75,27419
Nop-11	40	24,005	25	86,56667
Des-11	30	23,90323	28	87,69355
Jan-12	310	23,96613	28	83,57258
Feb-12	224	23,72414	20	84,37069
Mar-12	178	23,85806	26	82,26613
Apr-12	121	24,3	16	82,49167
Mei-12	143	23,8129	11	81,82258
Jun-12	109	23,045	5	78,94167
Jul-12	125	22,17258	3	78,66935
Agust-12	59	22,39355	2	74,74194
Sep-12	46	23,86833	2	71,01667
Okt-12	41	24,94435	10	73,92742
Nop-12	49	24,66167	22	82,98333
Des-12	46	23,98065	30	87,12903
Jan-13	262	24,11774	29	85,6371
Feb-13	216	24,0875	23	85,11607

Mar-13	212	24,13871	20	85,14516
Apr-13	104	24,52833	18	85,61667
Mei-13	140	24,45968	17	83,62903
Jun-13	126	26,22	11	84,68333
Jul-13	145	22,64758	8	84,54839
Agust-13	69	22,66613	2	77,79839
Sep-13	53	23,54	0	73,45833
Okt-13	49	24,70968	11	75,26613
Nop-13	59	23,9	19	83,46667
Des-13	63	23,38871	27	88,1371
Jan-14	84	23,79677	19	84,35484
Feb-14	77	23,98929	19	82,53571
Mar-14	65	24,05161	10	84,06452
Apr-14	83	23,95083	14	84,75
Mei-14	83	24,64839	6	82,30645
Jun-14	77	24,10333	5	82,51667
Jul-14	62	22,85323	1	82,55645
Agust-14	77	22,74768	1	78,36688
Sep-14	45	23,74185	1	76,52962
Okt-14	60	25,01573	5	75,47492
Nop-14	59	24,49212	10	82,68903
Des-14	62	24,01458	13	86,61735

BAB VII

KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran terkait pengerjaan tugas akhir ini. Bagian kesimpulan berisi kesimpulan dari hasil pengerjaan tugas akhir ini. Sedangkan pada bagian saran, berisi saran dan masukan untuk perbaikan, tindak lanjut, atau pengembangan penelitian tugas akhir ini.

7.1 Kesimpulan

Berdasarkan proses penelitian yang telah dilakukan, berikut ini merupakan kesimpulan yang dapat diambil:

1. Model regresi sederhana yang melibatkan variabel dependen AK mempunyai hasil prediksi yang tidak begitu baik. Nilai MAPE pada data latih berkisar antara 61,07%-80,33%. Sedangkan nilai MAPE pada data uji berkisar antara 61,22%-79,38%. Model yang melibatkan curah hujan sebagai variabel independen merupakan model dengan prediksi terbaik berdasarkan data latih, sekaligus menjadi yang terburuk berdasarkan data uji.
2. Heteroskedastisitas dapat tereduksi dengan membuang penciran dari data observasi, walaupun tidak menjamin akan sepenuhnya menghilang. Pelanggaran terhadap asumsi ini menunjukkan kemungkinan hubungan kedua variabel yang tidak linear, misalnya polynomial, akar kuadrat, atau logaritmik.
3. Transformasi log pada variabel dependen memperbaiki asumsi model, namun tidak begitu berpengaruh pada hasil prediksi.

4. Model yang menggunakan variabel dependen berupa natural log AK menghasilkan prediksi dengan nilai MAPE yang baik, yakni berkisar antara 11,00%-13,67% untuk data latih dan berkisar antara 8,88%-11,89% untuk data uji. Model dengan variabel independen curah hujan kembali menjadi model dengan hasil prediksi terbaik berdasarkan data latih sekaligus terburuk berdasarkan data uji.
5. Variabel curah hujan lebih berpengaruh terhadap prediksi jumlah penderita demam berdarah daripada kelembaban dan temperatur. Variabel AK juga berkorelasi terbaik dengan variabel curah hujan. Sehingga baik model H2a dan H2b sama-sama mempunyai koefisien determinasi yang paling baik, yakni 0,4778 dan 0,4439.
6. Variabel temperatur dan kelembaban tidak cukup memberikan kontribusi informasi yang signifikan apabila sudah terdapat variabel curah hujan sebagai variabel independen dalam memprediksikan variabel AK. Ini juga berlaku untuk variabel natural log AK.
7. Tingkat signifikansi variabel temperatur dan curah hujan pada kelompok model A dan B adalah $<.001$. Sedangkan untuk variabel kelembaban berada pada level signifikansi $<.01$.
8. Model prediksi, khususnya regresi, yang dapat menggambarkan jumlah penderita demam berdarah di suatu daerah dengan memperhatikan faktor cuaca pada saat itu dapat bermacam-macam. Data masukan yang berbeda dapat menghasilkan model efisien yang berbeda pula, baik dari segi komposisi variabel bahkan pola hubungan antara variabel respon dan prediktor.
9. Regresi OLS cocok digunakan apabila variabel dependen dan independen memiliki varian yang sama,

atau setidaknya yang tidak berbeda jauh. Sehingga variabilitas dari variabel dependen dapat dijelaskan dengan baik oleh variabel independen. Hal inilah yang juga menyebabkan model dengan variabel dependen natural log AK mempunyai nilai MAPE yang kecil, di samping kelemahan MAPE yang bias terhadap nilai prediksi yang rendah.

7.2 Saran

Berikut ini merupakan saran untuk penelitian selanjutnya:

1. Parameter-parameter cuaca lain, misalnya suhu minimum dan maksimum, presipitasi kumulatif, jumlah hari dalam sebulan yang memiliki temperatur melebihi batas tertentu, dan parameter-parameter turunan yang lain, kiranya dapat membantu untuk menemukan kondisi terbaik yang memicu peningkatan jumlah penderita DBD. Selain itu, penambahan parameter berpeluang besar dalam meningkatkan koefisien determinasi model.
2. Ketersediaan data observasi yang lebih banyak diharapkan mampu menangkap faktor siklus lima tahunan pada kasus demam berdarah yang terjadi di Indonesia.
3. Referensi dari studi ekologi terkait demam berdarah dapat membantu memberikan hipotesis awal mengenai parameter-parameter yang mungkin berpengaruh terhadap angka kejadian demam berdarah pada suatu daerah.
4. Prediksi demam berdarah dapat dikembangkan menjadi prediksi kemunculan endemik demam berdarah di suatu lingkup daerah, misal di tingkat kecamatan.

5. Sebelum menentukan metode regresi yang akan digunakan sebaiknya terlebih dahulu menganalisis data-data masukan. Regresi OLS baik digunakan apabila peneliti telah mempunyai keyakinan bahwa prediktor yang digunakan diukur tanpa error (bebas error) dan varians antara prediktor dan variabel respon tidak jauh berbeda.
6. Ketimpangan varians variabel-variabel dalam model dapat diatasi dengan transformasi variabel respon atau prediktor. Ketika transformasi dilakukan, hubungan antara variabel respon dan prediktor tidak lagi linear.
7. Metode regresi lain yang dapat dicoba adalah Orthogonal Regression dan Geometric Mean Regression. Kedua metode regresi ini dapat digunakan ketika variabel dependen dan independen mempunyai kecenderungan data yang bersifat acak [28].

Daftar Pustaka

- [1] WHO, “Dengue and severe dengue,” 2015. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs117/en/>. [Diakses 6 Mei 2015].
- [2] viva.co.id, “21 Daerah di Jawa Timur Berstatus KLB Demam Berdarah,” 2015. [Online]. Available: <http://nasional.news.viva.co.id/news/read/584031-21-daerah-di-jawa-timur-berstatus-klb-demam-berdarah>. [Diakses 6 Mei 2015].
- [3] Departemen Kesehatan RI, “Tata Laksana DBD,” 2017. [Online]. Available: <https://id.scribd.com/doc/163579819/Tatalaksana-DBD-Depkes-2007>. [Diakses 30 September 2015].
- [4] D. Rogers, M. Shapiro, G. Brunet, J. Cohen, S. Connor, A. Diallo, E. W. K. Haidong, S. Hales, D. Hemming, I. Jeanne, M. Lafaye, Z. Mumba, N. Raholijao, F. Rakotomanana, H. Teka, J. Trtanj dan P. Whung, “Health and Climate - opportunities,” *Procedia Environmental Science*, vol. I, pp. 37-54, 2010.
- [5] L. A. McLay, E. L. Boone dan J. P. Brooks, “Analyzing the volume and nature of emergency medical calls during severe weather events using regression methodologies,” *Socio-Economic Planning Science*, vol. 46, pp. 55-66, 2012.
- [6] A. Ekstrom, L. Kurland, N. Farrokhnia, M. Castren dan M. Nordberg, “Forecasting Emergency Department Visits Using Internet Data,” *General Medicine/Original Research*, 2014.

- [7] E. Descloux, M. Mangeas, C. E. Menkes, M. Lengaigne, A. Leroy, T. Tehei, L. Guillaumot, M. Teurlai, A.-C. Gourinat, J. Benzler, A. Pfannstiel, J.-P. Grangeon, N. Degallier dan X. deL, "Climate-Based Models for Understanding and Forecasting Dengue Epidemics," Creative Commons Attribution License, 14 Februari 2012. [Online]. Available:
<http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0001470>. [Diakses 14 Oktober 2015].
- [8] M. Gharbi, P. Quenel, J. Gustave, S. L. R. G. Cassadou, L. Girdary dan L. Marrama, "Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors," BioMed Central Ltd, 9 Juni 2011. [Online]. Available:
<http://www.biomedcentral.com/1471-2334/11/166/>. [Diakses 14 Oktober 2015].
- [9] Pusat Data dan Surveilans Epidemiologi Kementerian Kesehatan RI, "Demam Berdarah Dengue," *Buletin Jendela Epidemiologi*, vol. 2, pp. 1-14, 2010.
- [10] T. Y. M. Wahyono, B. Haryanto, S. Mulyono dan A. Adiwibowo, "Faktor-faktor yang Berhubungan dengan Kejadian Demam Berdarah dan Upaya Penanggulangannya di Kecamatan Cimanggis, Depok, Jawa Barat," *Buletin Jendela Epidemiologi*, vol. 2, pp. 31-43, 2010.
- [11] R. Weron, "Review: Electricity price forecasting: A review of the state-of-the-art," *International Journal of Forecasting*, vol. 30, pp. 1030-1081, 2014.
- [12] J. B. J. Guerard, "Regression Analysis and Forecasting Models," dalam *Introduction to Financial Forecasting in Investment Analysis*, New York,

- Springer Science+Business Media, 2013, pp. 19-45.
- [13] Misbahuddin dan I. Hasan, *Analisis Data Penelitian dengan Statistik*, Jakarta: Bumi Aksara, 2004.
- [14] A. D. Hervino, "Analisis Regresi Berganda & Pengujian Asumsi OLS," [Online]. Available: www.atmajaya.ac.id/filecontent/ekonomi-Analisis_Regresi_Berganda.ppt. [Diakses 2015 Oktober 1].
- [15] S. H. Yulianingtyas, *Pembuatan Aplikasi Peramalan Berbasis Web Berdasarkan Hasil Analisis Pengaruh Harga Emas Terhadap Harga Minyak, Nilai Tukar Mata Uang dan Indeks SP*, Surabaya: ITS, 2015.
- [16] Konsultan Statistik, "Regresi Linear Berganda," [Online]. Available: <http://www.konsultanstatistik.com/2009/03/regresi-linear.html>. [Diakses 1 Oktober 2015].
- [17] S. Makridakis dan M. Hibon, "Evaluating Accuracy (or Error) Measures," INSEAD, Fontainebleau, 1995.
- [18] SAS, "Mean Squared Error," [Online]. Available: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_intromod_sect005.htm. [Diakses 1 Oktober 2015].
- [19] Vanguard Software Corporation, "Mean Absolute Percentage Error," [Online]. Available: <http://www.vanguardsw.com/business-forecasting-101/mean-absolute-percent-error-mape/>. [Diakses 30 September 2015].
- [20] iPredict, "Time-series Forecasting Error Statistics," 20 Februari 2012. [Online]. Available: <http://www.ipredict.it/ErrorStatistics.aspx>. [Diakses 1 Oktober 2015].

- [21] J. S. Armstrong dan F. Collopy, "Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons," *International Journal of Forecasting*, vol. 8, pp. 69-80, 1992.
- [22] Microsoft, "Method to calculate interpolation step value in Excel," 23 September 2015. [Online]. Available: <https://support.microsoft.com/en-us/kb/214096>. [Diakses 5 Desember 2015].
- [23] C. Chatfield, *Time Series Forecasting*, Florida: CRC Press, 2000.
- [24] National Institute of Standards and Technology, "NIST/SEMATECH e-Handbook of Statistical Methods," 30 Oktober 2013. [Online]. Available: <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc435.htm>. [Diakses 18 Februari 2016].
- [25] J. J. Faraway, "Practical Regression and Anova Using R," 2002.
- [26] S. Prabhakaran, "How to detect heteroscedasticity and rectify it," 13 Januari 2016. [Online]. Available: <http://www.r-bloggers.com/how-to-detect-heteroscedasticity-and-rectify-it/>. [Diakses 21 Januari 2016].
- [27] D. Ferrari dan T. Head, "Regression in R Part I: Simple Linear Regression," 2010.
- [28] L. Leng, T. Zhang, L. Kleinman dan W. Zhu, "Ordinary Least Square Regression, Orthogonal Regression, Geometric Mean Regression and their Applications in Aerosol Science," *Journal of Physics: Conference Series*, pp. 78-82, 2007.
- [29] J. M. Brunkard, E. Cifuentes dan S. J. Rothenberg, "Assessing the roles of temperature, precipitation, and ENSO in dengue re-emergence on the Texas-Mexico border region," *Salud Publica Mex*, vol.

50, pp. 227-234, 2008.

- [30] J. Oh dan B. Kim, "Prediction Model for Demands of the Health Meteorological Information Using a Decision Tree Method," *Asian Nursing Research*, vol. IV, no. 3, pp. 151-162, 2010.
- [31] Prenhall, "Forecast Accuracy," [Online]. Available: <http://www.prenhall.com/divisions/bp/app/russelld/PROTECT/CHAPTERS/CHAP10/HEAD04.HTM>. [Diakses 30 September 2015].

Halaman ini sengaja dikosongkan.

BIODATA PENULIS



Lahir di Sumenep, Jawa Timur pada 20 Januari 1994, Rafika Nurmasari merupakan anak kedua dari tiga bersaudara dari pasangan suami istri Samsul Bahri dan Nurjannah. Masuk di Jurusan Sistem Informasi Institut Teknologi Sepuluh Nopember melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) Undangan pada tahun 2012.

Mata kuliah favorit penulis adalah Matematika Diskrit, Riset Operasi, Sistem Cerdas, Kecerdasan Bisnis, dan beberapa mata kuliah lain yang mempelajari tentang pengolahan data untuk mendukung pembuatan keputusan. Sehingga penulis memilih bidang minat laboratorium Rekayasa Data dan Inteligensi Bisnis sebagai laboratorium peminatan untuk tugas akhirnya.

Penulis menikmati kegiatan-kegiatan lain di luar perkuliahan, seperti mengikuti seminar dan workshop, kompetisi akademis, program mahasiswa wirausaha, pengajar sukarela, asisten dosen, serta kepanitiaan beberapa acara di lingkup jurusan hingga institut. Salah satu harapan besar penulis pasca kampus adalah menjadi lulusan ITS yang berkompeten dan dapat melanjutkan hobi mengajar yang dimilikinya.